STRUCTURAL INVESTIGATION OF THE BACTERIAL AMYLOID ADHESIN CURLI BY SOLID STATE NMR

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von Tobias Schubeis
aus Essen

1. Referentin: Professorin Dr. Christiane Ritter
2. Referent: Professor Dr. Michael Steinert
eingereicht am: 05.06.2013
mündliche Prüfung (Disputation) am: 02.12.2013

Druckjahr: 2014

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch die Mentorin der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Publikationen:

Schubeis, T, Lührs, T., Ritter C.: Unambiguous assignment of short and long range structural restraints by solid state NMR using segmental isotope labeling. DOI:10.1002/cbic.201402446 .

Tagungsbeiträge

Schubeis, T. & Ritter, C.: Segmental isotope labeling of insolube proteins. (Poster) EUROMAR Konferenz, Frankfurt am Main (2011)

CONTENTS

ABBREVIATIONS

Å	Ångström
AFM	Atomic force microscopy
AP	Alkaline phosphatase
APS	Ammonium persulphate
BMRB	Biological magnetic resonance data bank
BCIP	5-bromo-4-chloro-3-indolylphosphate
Bis-Tris	Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methan
CD	Circular dichroism
CP	Cross Polarization
CR	Congo red
CsgX	Genproduct of curli specific gene X
CV	Column volume
Da	Dalton
DARR	Dipolar assisted rotational resonance
DMSO	Dimethyl sulfoxide
E. coli	*Escherichia coli*
EM	Electron microscopy
ECL	Electrochemiluminescence
EDTA	Ethylenediaminetetraacetic acid
FA	Formic acid
GuHCl	Guanidine Hydrochloride
H/D	Hydrogen/Deuterium
HSQC	Heteronuclear single quantum coherence
FTIR	Fourier transform infrared
HRP	Horse radish peroxidase
IB	Inclusion body
IMAC	Immobilized metal ion affinity chromatography
IPTG	Isopropyl--D-thiolgalactoside
KPi	KH2PO4/K2HPO4
MWCO	Molecular weight cut off
Ni-NTA	Nickel-nitrilotriacetic acid-agarose
NMR	Nuclear magnetic resonance
PAGE	Polyacrylamid gel electrophoresis
PAR	Proton assisted recoupling
PCR	Polymerase chain reaction
PDB	Protein data bank
PDSD	Proton driven spin diffusion
R1-5	CsgA Repeat 1-5

SDS	Sodium dodecyl sulphate
Sec	Secretory
ssNMR	Solid state nuclear magnetic resonance
TAE	Tris-acetate-EDTA
TBS	Tris buffered saline
TCA	Trichloroacetic acid
TCEP	Tris(2-carboxyethyl)phosphine
TEM	Transmission electron microscopy
TEMED	N, N, N', N' – Tetramethylendiamin
TFA	Trifluoroacetic acid
ThT	Thioflavin T
TPPM	Two pulse phase modulation
Tris	Tris(hydroxymethyl)aminomethane
wt	Wild type

Amino acids are abbreviated using 1 or 3 letter code.
Latinisms are written in italic

ZUSAMMENFASSUNG

CsgA ist der Hauptbestandteil eines bakteriellen funktionellen Amyloids, genannt Curli, welches von Enterobakterien wie z.B. *Escherichia coli* exprimiert wird. Curli ist an der Bildung von Biofilmen und der Bindung an Wirtszellen beteiligt. CsgA amyloide wurden *in vivo* als auch *in vitro* ausgiebig untersucht und bilden ein geeignetes System um den Prozess der Amyloidbildung, Ausbreitung und deren Steuermechanismen zu studieren und zu verstehen. Hochaufgelöste Strukturen von Amyloiden sind die Voraussetzung um Einblicke in deren Funktion zu erlangen und vor allem um die Unterschiede zwischen krankheitsbedingten und funktionellen Amyloiden zu verstehen. Dazu wurden CsgA Fibrillen mittels Festkörper-NMR untersucht. Spektren von vollmarkierten Proben (^{13}C, ^{15}N) zeigten eine starke Resonanzüberlagerung aufgrund der intrinsischen Sequenz- und Strukturwiederholungen. Daher wurde eine Methode verwendet in der nur ein bestimmtes Teilstück des Proteins mit Isotopen markiert wird. Erreicht wird dies mit Hilfe von Intein katalysiertem Protein Spleißen. Dazu wurden verschiedene Intein-Fusionsproteine generiert um Proben mit unterschiedlichen Markierungsmustern zu erzeugen. Festkörper-NMR Messungen auf einer Probe mit einer einzelnen markierten strukturellen Wiederholung ermöglichten Sequenz spezifische Resonanzzuordnungen und erste Einblicke in die Struktur von CsgA. Das vermutete Strang-Schleife-Strang Motiv konnte bestätigt werden und Messungen von Abstandsbeschränkungen führten zu der Vermutung eines β-Strang-Austausches zwischen zwei aufeinanderfolgenden Molekülen innerhalb der Amyloidfibrillen.

Die für CsgA etablierte Strategie der segmentalen Isotopenmarkierung wurde außerdem auf die strukturell gut charakterisierten Amyloidfibrillen von HET-s (218-289) angewandt. Die erhaltenen Spektren zeigten keine Abweichung der chemischen Verschiebung, weder im Vergleich zu einer vollmarkierten Probe noch zu publizierten Daten. Das Ebenen spezifische Markierungsmuster ermöglichte außerdem die Messung von eindeutig intramolekularen Abständen.

SUMMARY

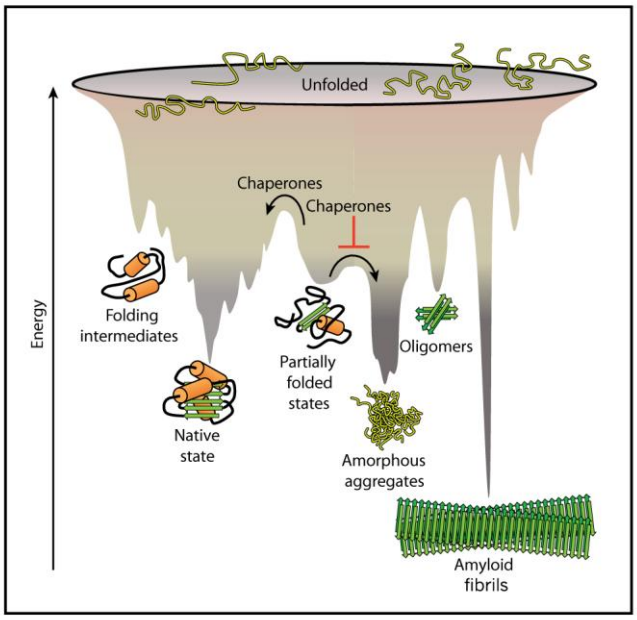
CsgA is the major subunit of a bacterial functional amyloid called curli, expressed by enterobacteriaceae e.g. *Escherichia coli*. Curli is involved in formation of biofilms and host cell attachment. Amyloids of CsgA have been extensively studied *in vivo* as well as *in vitro* and provide a well-suited system for understanding the process of amyloid formation, propagation and control mechanism. High-resolution structure of functional amyloid fibrils is a prerequisite to gain insights into their functions, mechanism, and more importantly the differences between disease-related and functional amyloids. Solid state NMR was used to investigate CsgA amyloid fibrils. Since spectra of the uniformly labeled CsgA suffered from high spectral overlap due to the high degree of sequence and structure homology in CsgA, segmental labeling by protein-trans-splicing was utilized. Intein fusion pairs for the preparation of three differently labeled samples were established. Solid state NMR measurements on a sample with single labeled structural repeat enabled sequence specific assignments and first insights to the structure of CsgA. The proposed strand-loop-strand motive could be confirmed and distance restraints gave rise to a β-strand swap within the amyloid fibril.

The established strategy of segmental isotope labeling was additionally demonstrated on the structurally well characterized amyloid fibrils of HET-s (218-289). The obtained spectra did not show chemical shift deviations from a uniformly labeled sample and published data. The layer specific labeling pattern allowed the detection on unambiguously intermolecular distance restraints.

1 INTRODUCTION

1.1 PROTEINS AND AMYLOIDS

The two distinctive parts of every living organism are deoxyribonucleic acid (DNA) and proteins. DNA is the genetic material consistent from one cell to the other. It contains a succession of four nucleotides that, according to the genetic code, is translated into amino acid sequences which fold to form proteins. Proteins are the active components of the cell: they are involved in almost every process, like biochemical catalysis, transport, storage, signal processing and DNA transcription regulation. The study of protein functions and interactions is a key factor for understanding cellular mechanisms. This understanding is of great importance to gain insight about, for example, the causes of diseases (Berg JM 2002).



.

Figure 1: Funnel shaped energy landscape of protein folding. Unfolded proteins are influenced by intra- and intermolecular interactions, e.g. by chaperones. They can adopt several stages of energy minima while the native fold is usually a global minimum. Other folds, like the formation of amyloid fibrils, can even have a lower energy state (reproduced from (Hartl, Bracher et al. 2011) with permission of the Nature Publishing group).

Proteins are composed of amino acids, which have a common building block based on the α carbon and a side chain, linked together through an amide bond. The differences between the 20 amino acids existing in nature are found in the side chains, which vary in size, charge and hydrophobicity. The amino acid chain can fold spontaneously into a stable structure with a minimum free energy. The folding occurs through the formation of hydrogen bonds, ionic and hydrophobic interactions between amino acids. The amino acid sequence, also called the primary structure, adopts, upon folding, secondary structure elements named α helix and β sheets (Anfinsen 1973). Specifically, in α helices every backbone amide proton is involved in a hydrogen bond to the carbonyl group of the amino acid four residues earlier in the sequence

resulting in a right-handed coiled or spiral conformation (Pauling, Corey et al. 1951). In β sheets, several amino acid stretches align parallel or antiparallel to each other. These β strands are connected laterally by at least two or three backbone hydrogen bonds resulting in a pleated sheet (Pauling and Corey 1951). The arrangement of such elements leads to the so-called tertiary structure of proteins, which represents the native and functional state of the protein but is difficult to predict from the primary sequence (Ivankov and Finkelstein 2004).

Considering the widely accepted funnel like energy landscape for protein folding, as illustrated in figure 1, it becomes obvious that misfolding, aggregation and even alternative stable conformations are possible (Dill and Chan 1997). Misfolding can occur e.g. by ribosomal translation failures, incorrect assembly of secondary structure elements and exposure of hydrophobic residues resulting in aggregation prone intermediates (Herczenik and Gebbink 2008, Hartl and Hayer-Hartl 2009). Protein aggregation can for instance result in highly disordered amorphous aggregates or in highly ordered robust and thermodynamically stable fibrillar arrangements: the so-called amyloid fibrils (Bonar, Cohen et al. 1969, Glenner, Cuatrecasas et al. 1969, Sipe and Cohen 2000). Protein misfolding takes place *in vivo* to a certain extend but it is under tight control of chaperons and the degradation machinery of the cell (Goldberg 2003). However, a rising number of protein misfolding diseases has been discovered, in which certain proteins overcome the control mechanism and form large aggregates. Amyloid depositions have been found to be the cause of many neurodegenerative diseases, such as Alzheimer's disease and prion diseases, in which amyloid accumulation leads to cell death in the brain. An overview of the most important disease related amyloids is given in table 1 (Chiti and Dobson 2006, Uversky and Fink 2006, Eisenberg and Jucker 2012).

Transmissible spongiform encephalopathies are caused by misfolding of the cellular prion protein (Prp^{C}) into an infectious form (Prp^{Sc}). The prion protein in its misfolded form is an infective agent (Prusiner 1982). The conformational change of a prion is self propagating, meaning that misfolded proteins are able to induce misfolding of surrounding prion proteins. This self-templating or seeding activity forms the basis for infectivity (Aguzzi, Sigurdson et al. 2008).

Table 1: Some amyloid forming proteins and associated human diseases

Disease	Associated Protein	Native Fold	Number of residues
Alzheimer's disease	Amyloid β	unfolded	40-42
Parkinson's disease	α-Synuclein	unfolded	140
Huntington's disease	Huntingtin Poly Q	unfolded	3144
Amyotrophic lateral sclerosis	Superoxide dismutase I	All-β	153
Spongiform encephalopathies	Prion Protein	unfolded / α-helical	253
Type II Diabetes	Amylin	unfolded	37

In addition, several living organisms take advantage of the ability of proteins to form amyloid fibrils to generate diverse biological functions (Fowler, Koulov et al. 2007, Hammer, Wang et al. 2008). In contrast to the disease related proteins, the amyloid fold is their native structure and not associated with protein misfolding. Functional amyloids have first been found in bacteria, yeast and fungi but have now been described in species from all kingdoms (Wickner, Edskes et al. 1999, Fowler, Koulov et al. 2006). Some functional amyloids are represented in table 2. Curli is one of the best studied functional amyloids and will be described in more detail in the following section. The morphology and biophysical characteristics of all amyloid fibrils are remarkably similar although the protein primary structure differs strongly among amyloid forming proteins (Ross, Edskes et al. 2005).

Table 2: Some functional Amyloids.

Amyloid Protein	Organism	Amyloid Function
Curli	*Escherichia Coli*	Biofilm component, adhesion to surfaces
FapC	*Pseudomonas*	Biofilm formation
Mtp	*Mycobacterium tuberculosis*	Pili formation, binding to laminin
Chaplins	*Streptomyces sp.*	Spore surface protein
HET-s	*Podospora anserine*	Regulation of heterokaryon formation
Ure2p	*Saccharomyces cerevisiae*	Regulation of nitrogen metabolism

As shown by electron microscopy in Fig. 2a, amyloids are fibrous protein multimers (Shirahama and Cohen 1965). Such fibrils have diameters ranging from 3 to 12 nm and varying length. The high content of β sheets can simply be detected by CD spectroscopy: a typical CD spectrum is depicted in Fig. 2e. Each monomer contributes a single or multiple layers of β– strands which arrange perpendicular to the fibril axis as illustrated in Figure 2b. This consistent array gives rise to a characteristic X-ray diffraction pattern (Eanes and Glenner 1968). Figure 2c shows the amyloid typical diffraction lines corresponding to interstrand (4.7 Å) and stacking distances (10 Å). Due to this cross like pattern, the amyloid characteristic structural element is called “cross β structure”. This structural motif is also detectable by a characteristic wavenumber maximum (approximately 1620 cm^{-1}) in Fourier transform infrared (FTIR) spectroscopy, as highlighted in Fig. 2 d (Termine, Eanes et al. 1972, Caughey, Dong et al. 1991).

Another finger print property of amyloids is the binding of the small molecular dyes Congo Red and Thioflavin T (Vassar and Culling 1959, Howie and Brewer 2009). The detectable effects upon binding are Birefringence of Congo Red and a shift of the absorption maximum of Thioflavin T. The latter dye is commonly used to detect amyloid aggregation kinetics *in vitro*, as depicted in Fig. 2f (LeVine 1999). The limiting step in the process is the formation of nuclei or seeds in the so called lag-phase where no significant signal increase is detectable. The length of the lag-phase is protein depended and varies from minutes and several days. Once induced, the amyloid formation propagates exponentially resulting in a sigmoidal curve. The amyloid formation can be effectively accelerated by adding preformed seeds which can lead to higher

homogeneity and efficient templating of a given conformation (Jarrett and Lansbury 1993, Harper and Lansbury 1997).

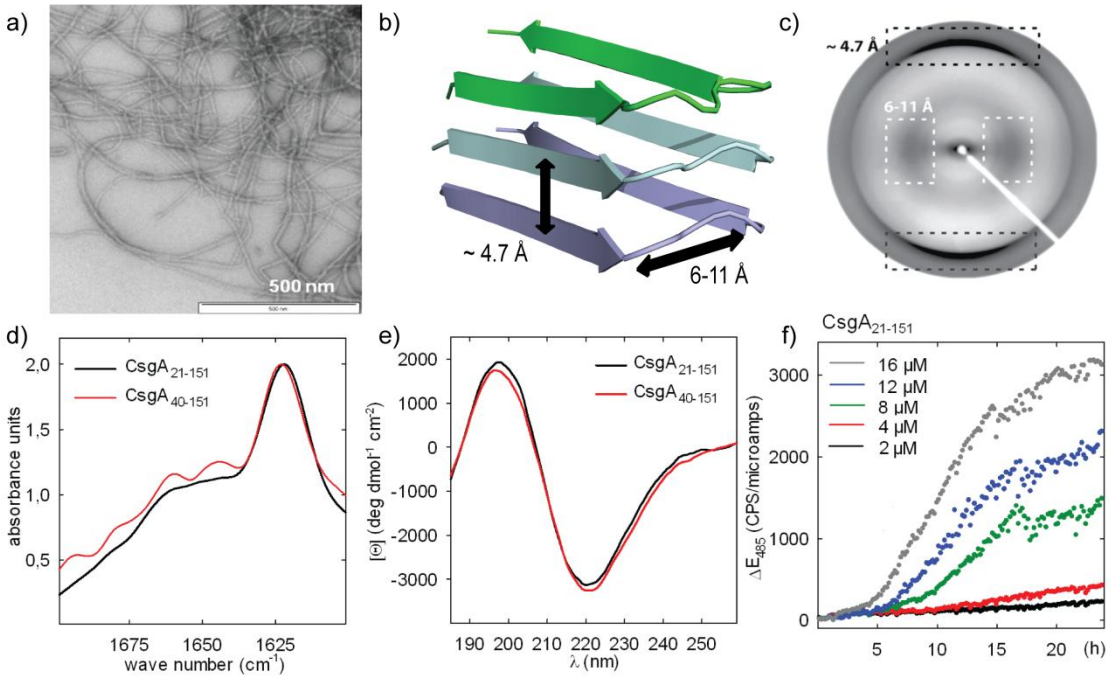


Figure 2: Amyloid characteristics. a) Electron micrograph of in vitro formed amyloid fibrils (reprinted with permission from M. Nagaraj), b) structural model indicating a cross β fold, c) X-ray diffraction pattern showing the distinctive diffraction lines at 10 and 4.7 Å, reproduced from Greenwald and Riek (2010) with permission from Elsevier, d) FTIR spectra of CsgA amyloid fibrils showing a maximum at 1620 cm-1, e) CD-spectra of CsgA amyloids showing an all-β secondary structure, f) in vitro aggregation assay followed by Thioflavin T fluorescence increase over the time (d/e/f reprinted with permission from A. Zimmer)

1.2 Curli Biogenesis

Curli are protein fibrils found in the extracellular matrix of many enterobacteria like *Escherichia coli* and *Salmonella*. In 1989 Olson and coworkers first described Curli as a novel class of cell adhesin from *E.coli* strains that caused bovine mastitis, but the proposal of Curli as functional amyloid was identified by Chapman and coworkers in 2002 (Olsen, Jonsson et al. 1989, Chapman, Robinson et al. 2002). Curli amyloids are part of biofilms, which are assemblages of bacterial cells attached to biotic or abiotic surfaces and enclosed in an adhesive matrix mostly composed of polysaccharides, extracellular DNA and proteins (Kikuchi, Mizunoe et al. 2005, Barnhart and Chapman 2006). The high resistance of Curli against protease digestion and denaturation helps the bacterial community to shield from harsh environmental stresses (Van Houdt and Michiels 2005). Recently, Curli have been shown to be involved in surface binding, host invasion and binding to host proteins like toll-like receptors, leading to activation of the innate immune system (Austin, Sanders et al. 1998, Gophna, Barlev et al. 2001, Tukel, Raffatellu et al. 2005, Tükel, Nishimori et al. 2010).

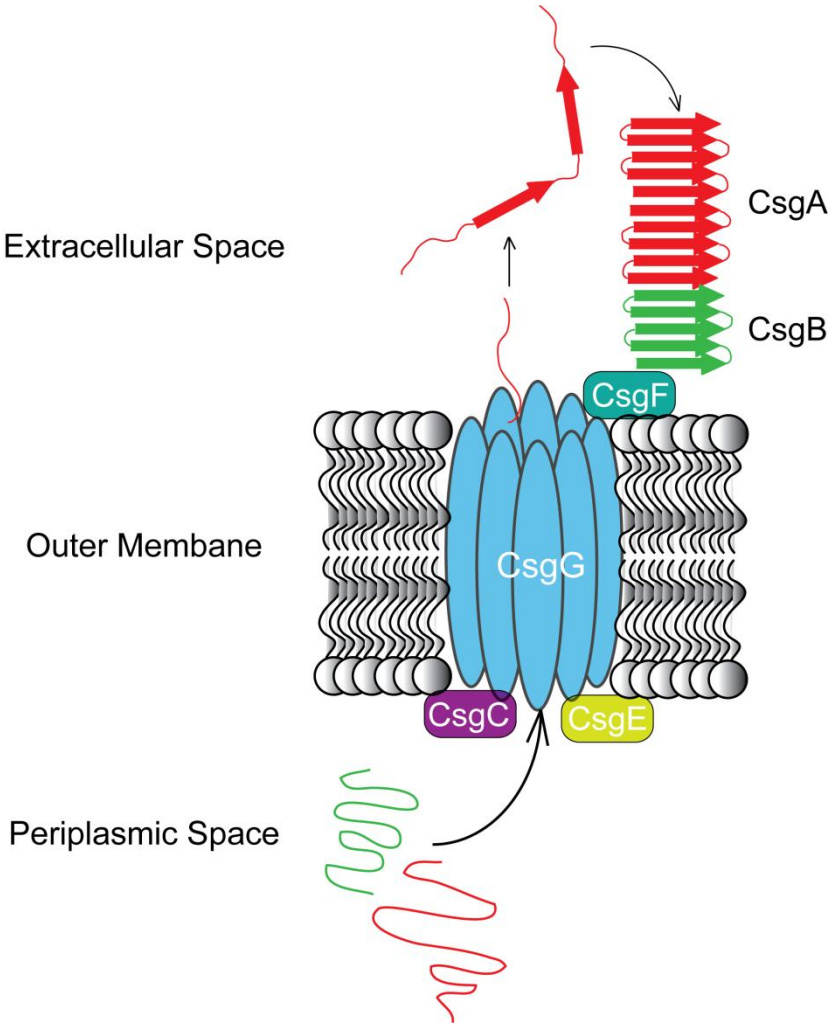


Figure 3: Curli Biogenesis. All Csg-proteins are translocated by the Sec-system to the periplasm. CsgG form a multimeric trans-membrane complex to export unfolded, monomeric CsgA and CsgB. CsgC and CsgE assist in the periplasm while CsgF is transported to the outer membrane surface. Once exported CsgB folds rapidly and nucleates the controlled aggregation of CsgA to form a fibril.

The curli fibril expression and formation of enterobacteria is highly regulated and involves two divergently expressed operons. A total number of seven proteins are involved in the Curli biogenesis. They are encoded in the csg genes (Hammar, Arnqvist et al. 1995, Romling, Bian et al. 1998). The csgBAC operon encodes for the major curli subunit CsgA, the nucleator protein CsgB and CsgC which is necessary for the correct assembly of curli fibrils. The csgDEFG operon encodes for CsgD - the transcription activator for the csgBAC operon, for the two chaperons CsgE and CsgF as well as for the transmembrane exporter complex CsgG (Gerstel, Park et al. 2003). Once expressed, all Csg - proteins get exported to the periplasm by the Sec System using a specific export recognition sequence at their N-terminus. CsgG forms a ring like hexameric to octameric outer membrane transporter which mediates the export of CsgA and CsgB (Robinson, Ashman et al. 2006). CsgE and CsgF are believed to act as chaperones to keep the major curli subunit and the nucleator in an unfolded state. CsgE binds CsgG in the periplasm while CsgF gets exported as well and facilitates the cell adhesion of CsgB (Nenninger, Robinson et al. 2009, Nenninger, Robinson et al. 2011). After secretion of CsgA and B, the polymerization of soluble CsgA monomers is nucleated by CsgB, which is attached

to the outer membrane, resulting in unbranched cell surface associated amyloid fibrils (Bian and Normark 1997, Chapman, Robinson et al. 2002). The Curli biogenesis is summarized in Figure 3.

Curli fibrils consist out of the major Curli subunit CsgA and the nucleator protein CsgB. Both proteins are able to form amyloid fibrils *in vitro*. They share 30% of sequence identity and 40% homology (Hammar, Arnqvist et al. 1995, Hammer, Schmidt et al. 2007). The primary sequence of CsgA contains the N-terminal Sec signal followed by 22 flexible residues and five imperfect repetitive units (R1-R2-R3-R4-R5). The Sec signal is cleaved off right after translocation to the periplasm. The 22 N-terminal residues are sensitive to proteinase K digestion, indicating surface exposition and flexibility (White, Collinson et al. 2001). Certain residues are proposed to be an export signal domain to the extracellular matrix (Robinson, Ashman et al. 2006). The five repeating units are protease resistant and form the amyloid core (Collinson, Parker et al. 1999). This core contains intramolecular highly conserved amino acids. Every repeat shows the sequence $Sx_5QxGxGNxAx_3Qx$ and consists of 19-23 residues. The central QxGxxN motive is furthermore conserved over a variety of species.

Recombinantly expressed and fibrilized CsgA shows the same amyloid characteristics like the wild type Curli purified from *E.coli* biofilms (Wang, Smith et al. 2007). They are highly stable and insoluble even under strong denaturing conditions (6M GuHCl, 0,1% SDS) (Collinson, Emody et al. 1991). Additionally, it has been shown that synthesized peptides of R1, R3 and R5 can form amyloid fibrils themselves while R2 and R4 peptides do not (Wang, Smith et al. 2007).

In a structural model of CsgA proposed by Collinson *et al.*, the glutamine and asparagine sidechains are suggested to form hydrogen bonds defining a stable strand-loop-strand structure with parallel β sheets perpendicular to the fibril axis (Collinson, Parker et al. 1999). In a solid state NMR study on CsgA, Shewmaker and coworkers concluded that CsgA fibrils are not based on parallel in-register β sheets. However, due to the experimental design of this study, they did not obtain high quality data sufficient to calculate a three dimensional structure of CsgA fibrils (Shewmaker, McGlinchey et al. 2009). In a previous H/D-exchange NMR based study from our group, it has been proven that the residues 21-43 are flexible and unstructured and the precise locations of β-strands have been identified (Zimmer 2011). Every repeat contains two β-strands with a length of 7-8 residues as illustrated in Figure 4.

Figure 4: Primary structure alignment of CsgA 43-152 as five imperfect repetitive units (R1-R5) and secondary structure element location based on H/D exchange analysis.

1.3 TECHNIQUES TO STUDY AMYLOIDS

Bacterial amyloid systems provide us with facilitated models to study mechanisms of amyloid formation, regulation, degradation and function. Insights into mechanisms and functional differences of disease related and functional amyloids rely on the elucidation of high resolution structures (reviews by: (Chiti and Dobson 2006, Toyama and Weissman 2010).

There are several possibilities how to organize β strands in fibrillar systems, all summarized in figure 5. In a parallel arrangement they are usually organized in register, meaning that the same residues of two monomers line up underneath each other. In an antiparallel arrangement, every second monomer has an inverse orientation to the first. If a single monomer contributes to more than one layer of the fibril, it is called a β-solenoid. The β-solenoid can appear as a β-helix where three β sheets form a triangular interface or as a β-roll where just two sheets form the interface.

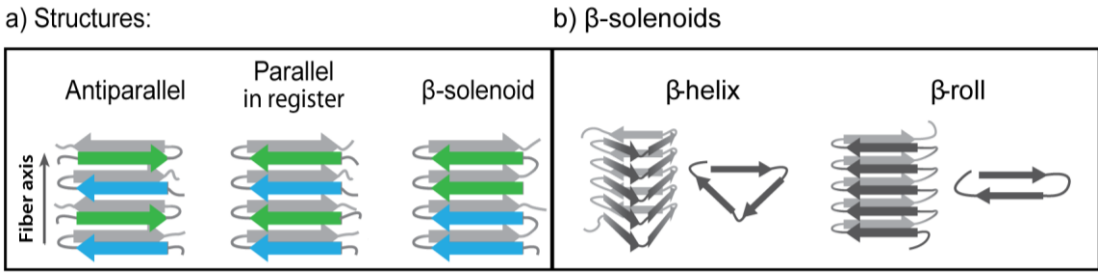


Figure 5: Classification of different amyloid folds. a) β-strands can either arrange in a parallel or antiparallel manner. If a single monomer contributes to more than one layer of β-strands in a fibril, the fold is called β-solenoid, b) schematic drawings of β-solenoids, in a β-helix three β-sheets form a triangular interface, in a β-roll two sheets face each other (reproduced from Toyama and Weissmann 2010 with permission from ANNUAL REVIEWS)

Due to their insolubility, large size, dynamic properties and intrinsic heterogeneity, amyloid fibrils are challenging targets for the most common techniques in structural biology, namely x-ray crystallography and solution nuclear magnetic resonance (NMR). For this reason, for many years only low resolution techniques were available to get insights to amyloid structures. As mentioned above, x-ray fibril diffraction and Fourier transform infrared spectroscopy can help identifying amyloid character of protein aggregates (Dueholm, Nielsen et al. 2011). Electron microscopy and atomic force microscopy can provide information about fibril morphology as well as ultrastructural characteristics, for example if fibrils branch or wind around each other (Knowles, Smith et al. 2006, Fitzpatrick, Debelouchina et al. 2013). Scanning transmission electron microscopy has been implemented as a method for mass-per-length determination and successfully been used on amyloid fibrils. This is particularly useful for the estimation of the number of monomers that contribute to each layer of the fibril (Chen, Thurber et al. 2009).

The primary sequence can be screened for positions that are of importance for amyloid formation through the introduction of point mutations which lead to the loss of secondary structure. For this purpose, prolines are suitable due to their structure disrupting properties that affect the β sheet organization (Chiba, Hagihara et al. 2003, Williams, Portelius et al. 2004).

Point mutations to cysteine bear the possibility to couple fluorophores, like pyrene, or to attach spin labels to specific positions. By this intra, as well as intermolecular distances between probes, can be measured either by fluorescence or electron paramagnetic resonance (EPR) spectroscopy (Torok, Milton et al. 2002, Krishnan and Lindquist 2005, Chen, Margittai et al. 2007). The prerequisite for a mutation and labeling study is that the modified protein shows the same fold and biochemical properties as those found for the wild type.

A more detailed structural insight can be obtained by Hydrogen-Deuterium exchange experiments. This method relies on the exchange of amide protons with the solvent (D_2O) deuterons. This effect is inhibited if the proton is involved in a hydrogen bond such as in secondary structural elements (Hvidt and Linderstrøm-Lang 1954, Englander and Kallenbach 1983). The exchange can be detected by mass spectrometry (MS) and solution nuclear magnetic resonance (Hoshino, Katou et al. 2002, Kheterpal, Cook et al. 2006). The analysis by MS is faster and easier but does not provide residue specific information whereas the NMR based method does. For this, a NMR backbone assignment is required for a precise investigation. In case of amyloids, several ^{15}N-labeled fibrillar samples have to be incubated for different times in D_2O before dissolving and measuring in anhydrous dimethyl sulfoxide. Two dimensional [^{15}N-^{1}H]-spectra are used to analyze peak intensity decrease over the time, making use of the different NMR properties of protons and deuterium. H/D-exchange experiments provide a powerful tool to identify structured regions on a residue specific level. Additional information about the sample stability and homogeneity can be evaluated (Luhrs, Ritter et al. 2005, Ritter, Maddelein et al. 2005). Also cryogenic electron microscopy has been used to investigate amyloid structures. It was possible to fit β sheet arrangements into the electron density and display the helical twist of fibrils but the resolution is not yet high enough to provide atomic details (Jimenez, Guijarro et al. 1999, Sachse, Fandrich et al. 2008). X-ray diffracting microcrystals of short peptides that formed amyloid-like fibrils have been obtained and enabled a detailed view of amyloid fibril structure (Nelson, Sawaya et al. 2005). Solid state NMR (ssNMR) has evolved to be the method of choice for high resolution structural investigations of non-crystalline solid materials like amyloid fibrils (Griffin 1998). Beside the secondary structure information, which can already be elucidated from the carbon backbone assignment by comparing the residue specific Cα/Cβ/CO chemical shifts to the average random coil values, also distances between nuclei can be analyzed and used for structure calculations (Castellani, van Rossum et al. 2002, Bockmann 2008, Comellas and Rienstra 2013). It has already been shown that ssNMR is capable to determine high resolution structures of microcrystalline proteins, membrane proteins as well as amyloid fibrils (Bockmann and Meier 2010, Hong, Zhang et al. 2012). The basics of solid state NMR will be discussed in more detail in section 1.5 and a more detailed description of structure calculation from NMR data is given in section 1.8.

1.4 AMYLOID STRUCTURES

The entire set of methods described above has been used to get structural insights to Alzheimer's disease (AD) causative agent, the Aβ peptide amyloids. Aβ is a 39-42 amino acid peptide, the proteolytic cleavage product of the amyloid β precursor protein (Kang, Lemaire et al. 1987). The most common isoforms are Aβ 1-40 and Aβ 1-42. The 40 residue peptide is the most abundant isoform in the brain while the 42 residue isoform is the most fibrillogenic (Näslund, Schierhorn et al. 1994) Even though the exact role of Aβ aggregates in AD is not yet understood, the structures of *in vitro* generated fibrils are of high interest. In an initial study on Aβ 34-42 peptides using a combination of ^{13}C-ssNMR and isotope edited FTIR. The peptides formed single straight β-strands which arrange antiparallelly in a fibril (Lansbury, Costa et al. 1995). Many other studies on truncated fragments of Aβ peptides led to the same conclusion. But investigations on the full Aβ 1-40 showed that those structures have an in register parallel β-sheet pattern (Antzutkin, Balbach et al. 2000, Petkova, Ishii et al. 2002). Finally, more and more studies concluded that the N-terminus is flexible and the C-terminus contains two to three β-strands (Whittemore, Mishra et al. 2005).

In a combinational approach of cryo EM, mutagenesis and H/D-exchange NMR, a Aβ 1-42 structure containing two β-strands with swapped domains was presented (Luhrs, Ritter et al. 2005). The structural model is shown in figure 6 a. Solid state NMR investigations using distant constraints concluded a comparable β sandwich structure for Aβ 1-40 (Petkova, Yau et al. 2006). Until today the secondary structure is of general agreement even though the exact positions of β-strands vary among the entries in the PDB. The tertiary structure and the exact arrangement of monomers in the fibril are still under debate. Additionally, Aβ is a system with a high structural diversity of the fibril albeit higher polymorphism was obtained for Aβ 1-40 than 1-42 (Schmidt, Sachse et al. 2009). At least two different models have been proposed based on solid state NMR and several more based on cryo EM studies (Paravastu, Leapman et al. 2008, Meinhardt, Sachse et al. 2009, Zhang, Hu et al. 2009). That is also one reason why Aβ fibrils do not serve as the perfect model fibrillar system for solid state NMR, since the occurrence of different folds in the sample leads to line broadening or to more than a single set of peaks for the same amino acid in the spectra (Lopez del Amo, Schmidt et al. 2012).

A much less diverse system is the HET-s amyloid. It was found in the filamentous fungus *Podospora anserina* where it has a role in controlled cell death, called heterokaryon incompatibility. The 289 amino acid protein can exist in a non-prion and in a prion form, but only the latter one is functional. Like every prion it is capable to convert non-prion HET-s proteins into the prion form, thereby amplifying its influence (Coustou, Deleu et al. 1997, Saupe 2007). Recombinant HET-s forms amyloid fibrils *in vitro* and shows activity in cell based assays. The prion domain has been identified by protease treatment and found to be located at the C-terminus. The residues 218-289 form a small (7kDa) protease resistant core. The N-terminal Domain has been crystallized and shows a globular, mainly α-helical fold (Dos Reis, Coulary-Salin et al. 2002, Maddelein, Dos Reis et al. 2002, Balguerie, Dos Reis et al. 2003, Greenwald,

Buhtz et al. 2010). A precise mapping of the β-strands in recombinant HET-s C-terminal domain was achieved by H/D exchange NMR and confirmed from a solid state NMR backbone assignment. Four β strands have been identified at residue numbers 226–234 (β1), 237–245 (β2), 262–270 (β3) and 273–282 (β4).

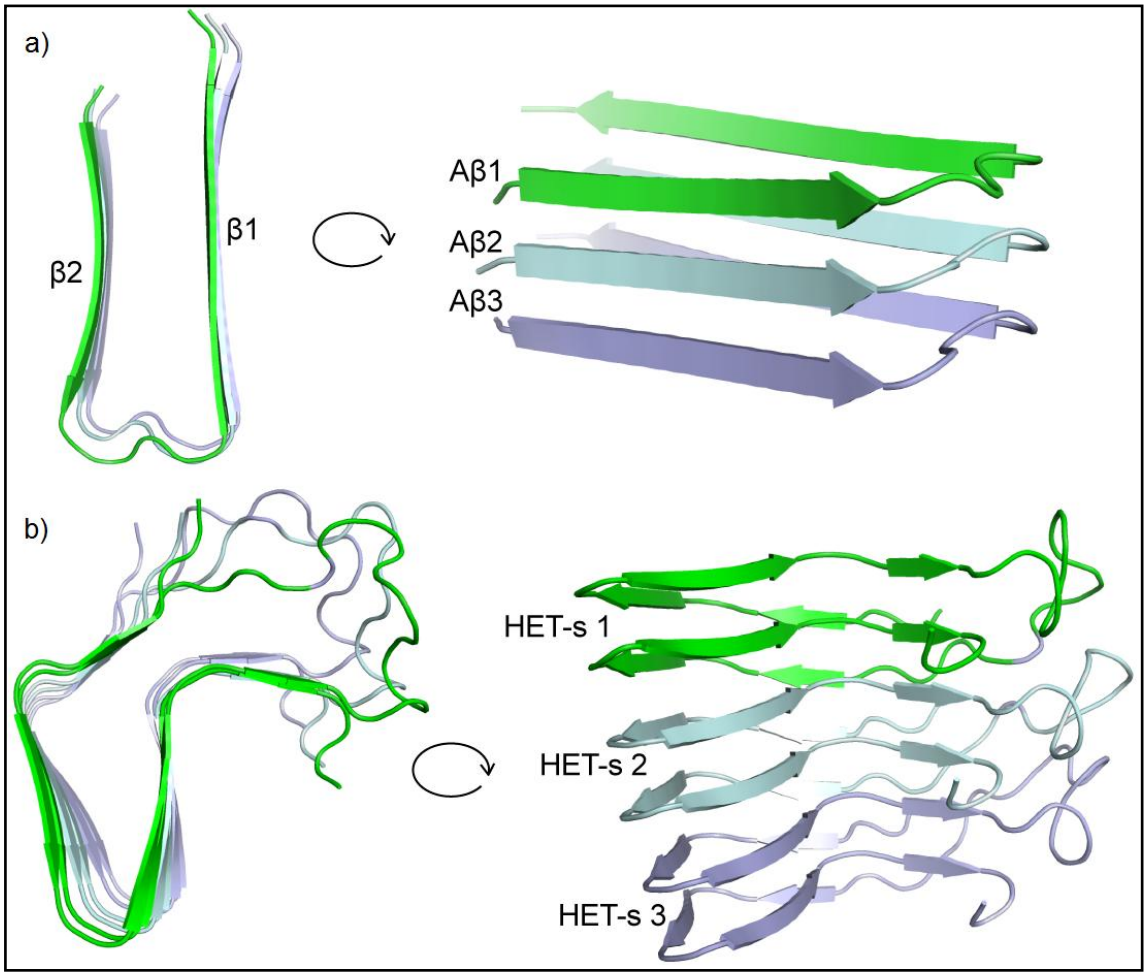


Figure 6: Model structure of Aβ1-42 drawn from the top and from the side (PDB ID: 2BEG), showing a parallel β-strand assembly and the HET-s(218-289) structure from the top and from the side (PDB ID 2RNM), showing a triangular β-solenoid.

A first model of a β roll in which β 1+3 and β 2+4 fold across from each other and form one layer of a fibril has been postulated. Neither the solid state spectra nor the H/D-exchange data did arise any suspicion of heterogeneity or multiple confirmations in the sample (Ritter, Maddelein et al. 2005).

Finally, an atomic resolution structure based on distance restraints from solid state NMR was published in 2008 (Wasmer, Lange et al. 2008). The required unambiguity was achieved using sparse isotope labeling and sample dilution. The high resolution structure revealed that each β-strand is split into two parts connected with a short β-arc. The ssNMR structure is shown in figure 6 b. The first three strands of every layer contribute to a triangular shape while the fourth strand is pointing away from the core. Most of the polar and charged amino acids are pointing out while the hydrophobic residues define the β solenoid fold. An even more refined structure, based on a higher number of distance restraints, has been published recently (Van Melckebeke,

Wasmer et al. 2010). The HET-s prion is so far the only high resolution solid state NMR structure and serves as a model system for amyloid structural investigations.

1.5 Structural Biology of rigid samples-an introduction to solid state NMR

The understanding of protein properties and functions relies on the elucidation of high resolution structures. X-ray crystallography and Nuclear Magnetic Resonance (NMR) Spectroscopy are well suited techniques to approach this objective: more than 98% percent of the structures deposited in the protein date base have been solved by one of the methods. High resolution X-ray crystallography requires well-ordered 3D protein crystals which diffract the x-ray beam. Growing these crystals is the major challenge of this technique.

NMR Spectroscopy in solution is a tool to study not only the structure but also the dynamics of proteins. However, due to physical properties, structure determination of proteins by NMR in solution is limited to a protein size of around 40 kDa. Membrane proteins in their native lipid environment and proteins with a tendency to oligomerize, like amyloid fibrils, are difficult to access by X-ray crystallography or NMR in solution. Membrane proteins often function as receptors and play critical roles in signal transduction. Mutations in membrane proteins as well as protein aggregation into amyloid fibrils are involved in several human diseases; therefore, the elucidation of high resolved structures is of major interest. Due to access to high field spectrometers and the advances in the development of methodology and isotope labeling techniques, solid state NMR has evolved to a powerful tool to study structures of insoluble macromolecules. The basics of NMR spectroscopy of rigid samples will be described in the following passage. A more detailed description of NMR theory is available in several textbooks (Wüthrich 1986, Cavanagh 2007, Levitt 2008).

Nuclei have an intrinsic angular momentum which is called the nuclear spin. NMR Spectroscopy relies on the interaction of nuclear spins in a static magnetic field (Bloch 1946). All nuclei with a spin quantum number unequal to zero are NMR active, but only spin ½ nuclei are relevant for biomolecular research (see table 3).

Table 3: A selection of nuclei, relevant for biomolecular NMR, and their properties.

Isotope	Spin	Relative abundance (%)	Gyromagnetic ratio (10^6 rad s^{-1} T^{-1})
^{1}H	½	100	267.53
^{2}H	1	0	41.07
^{12}C	0	98.9	%
^{13}C	½	1.1	67.28
^{14}N	1	99.63	19.34
^{15}N	½	0.37	-27.12

Among all nuclei, protons (^{1}H) have the largest gyromagnetic ratio and the highest natural abundance. Therefore, ^{1}H has the highest sensitivity in NMR experiments. Due to a relative abundance of around 1% for ^{13}C and ^{15}N, these isotopes have to be incorporated into protein

samples. For spin ½, each nucleus of a molecule can have two possible orientations, parallel or antiparallel, to an applied magnetic field (Zeeman 1897).

The energy difference between these states is given by:

$$\Delta E = \frac{\gamma h B_0}{2\pi}$$

Where γ is the gyromagnetic ratio, *h* is Plank's Constant and B_0 the external magnetic field. This means that the populations of the energy states depend on the type of nucleus and the external magnetic field. With higher external fields, the differences between nuclear spin energy levels become larger and the population differences between the states increases. The very small population majority of the lower energetic state gives rise to a bulk magnetization of the sample parallel to the magnetic field, which is defined as z-magnetization. In a thermodynamic equilibrium the spins precess around B_0 with the Larmor frequency $\omega = \gamma B_0$. If a radio frequency pulse with a frequency close to the Larmor frequency is applied from a position perpendicular to B_0, the thermodynamic equilibrium gets disturbed and the bulk magnetization changes orientation. For example, a pulse which turns the z-magnetization to the xy-plane is called a 90° pulse. After switching off the radio frequency pulse, the thermodynamic equilibrium is governed under the influence of two effects: the spin-lattice (T_1) and the spin-spin (T_2) relaxations. T_1 relaxation is caused by heat exchange with the lattice while T_2 relaxation occurs by interaction with surrounded nuclei (Bloembergen, Purcell et al. 1947). The relaxation results in a decay of the x, y magnetization. This is the so called "free induction decay" (FID) which is recorded by the NMR spectrometer. A basic NMR experiment is illustrated in figure 7.

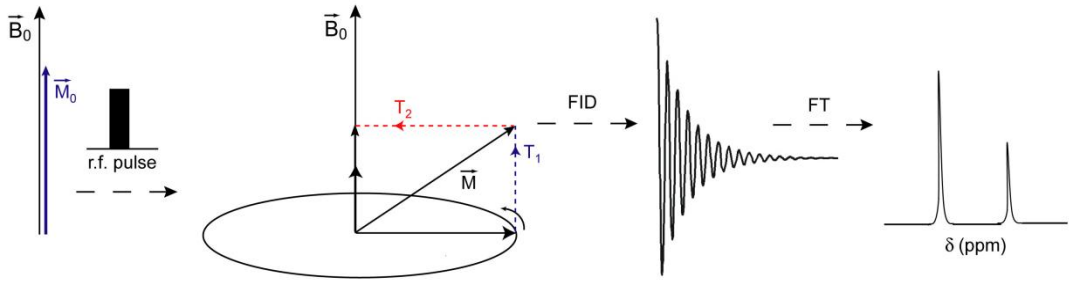


Figure 7: Pulsed Fourier transform NMR experiment. At thermal equilibrium the magnetization vector (M_0) is parallel to the magnetic field vector (B_0). By applying a radio frequency pulse, the magnetization is flipped and precesses in the XY-plane. The longitudinal magnetization is recovered with the relaxation time T_1 while the transverse magnetization decays with the relaxation time T_2. This free induction decay is recorded and Fourier transformed into the NMR spectra.

In modern NMR experiments radio frequency pulses of defined length, amplitude, phase and offset frequency, separated by signal evolution steps are applied according to a pulse program. The FID is finally Fourier transformed from a time (ms) to frequency domain (Hz) and the resonance frequencies are plotted relative to a standard. The difference to the standard is called chemical shift, which is usually very small and given in parts-per-million (ppm). The position and number of chemical shifts of nuclei are characteristic for the structure of a molecule. Commonly used reference substances are Tetramethylsilane (TMS) or 4,4-dimethyl-

4-silapentane-1-sulfonic acid (DSS) (Markley, Bax et al. 1998). During a NMR experiment each nucleus is influenced by several factors resulting in the final chemical shift in the NMR spectrum, these parameters are described in the following section.

1.6 NMR INTERACTIONS

All relevant physical interactions in a NMR experiment are quantum mechanically described by the Hamilton operator *H*, in which every possible interaction is described by a specific term (Mehring 1983). The calculation of *H* is made through the following equation; here, only the interactions of spin ½ nuclei are considered.

$$H = H_Z + H_{Rf} + H_{CS} + H_J + H_D$$

where:

- H_Z is the Zeeman interaction which is directly reliant on the Larmor Frequency ω_0 in the absence of any other interaction. It describes the magnetization along the z-axis in the static field. In case of a rotating reference frame with the frequency equal to ω_0, the term is eliminated from the equation.
- H_{Rf} describes the influence of magnetic components from the induced radio frequency field (rf) B_1. Compared to $B_{0,}$ this rf-field is very weak. State transitions of an observed nucleus can just be induced with frequencies close to the corresponding Larmor frequency.
- H_{CS} describes the chemical shift Hamiltonian, which is dependent on the surrounding electron density, the electronegativity of neighboring groups and anisotropic effects. The electron density shields the nucleus from the external magnetic field. Electron donating surrounding groups increase shielding while groups with a high electronegativity deshield the nucleus. The electronic environment of a nucleus is usually anisotropic, meaning it varies on different sides of a nucleus. In the quantum mechanical descriptions those anisotropic factors are introduced by a term $(3 \cos^2 \theta\text{-}1)$.
- H_J describes the J-coupling which is an electron mediated scalar interaction between two nuclei. J-couplings are particularly important for NMR of small molecules as well as for “through bond” transfer experiments in protein solution NMR.
- H_D describes the dipolar interactions, which are the most valuable coupling for solid state NMR. Nuclear spins exhibit a dipolar moment which interacts with dipolar moments of surrounded nuclei; this can be either homonuclear or heteronuclear. This interaction includes internuclear distance information. Like the chemical shift, the dipolar coupling depends on the orientation in the static magnetic field. This anisotropic effect leads to slight alterations of the chemical shifts as soon as the molecule changes the orientation in the magnetic field. While in solution NMR this effect is averaged out by the Brownian motion and only the isotropic chemical shift value is detected, this is not the case in solid state NMR. The following section will describe how the important isotropic information can be extracted in solid state NMR.

1.7 BASIC SOLID STATE NMR TECHNIQUES

1.7.1 MAGIC ANGLE SPINNING

Every NMR interaction with an orientation dependency and therefore an anisotropic effect in static samples leads to line broadening and gives rise to the so called powder pattern. In quantum mechanical equations the anisotropic influences are introduced by a term ($3\cos^2\theta$-1). The dependence on the angle θ between the internuclear vector and B_0 is exploited in solid state NMR by a technique named magic angle spinning (MAS). At an angle of 54.7° the term ($3\cos^2\theta$-1) becomes zero. By spinning the sample in an orientation of the magic angle to the external magnetic field with a rotor frequency higher than the anisotropic tensors, these interactions are averaged out over the NMR time-scale (Andrew 1981). Samples are measured in ZrO cylinders (Rotors) whose spinning is induced by an air stream along the cogwheel on the rotor cap like illustrated in figure 8 a.

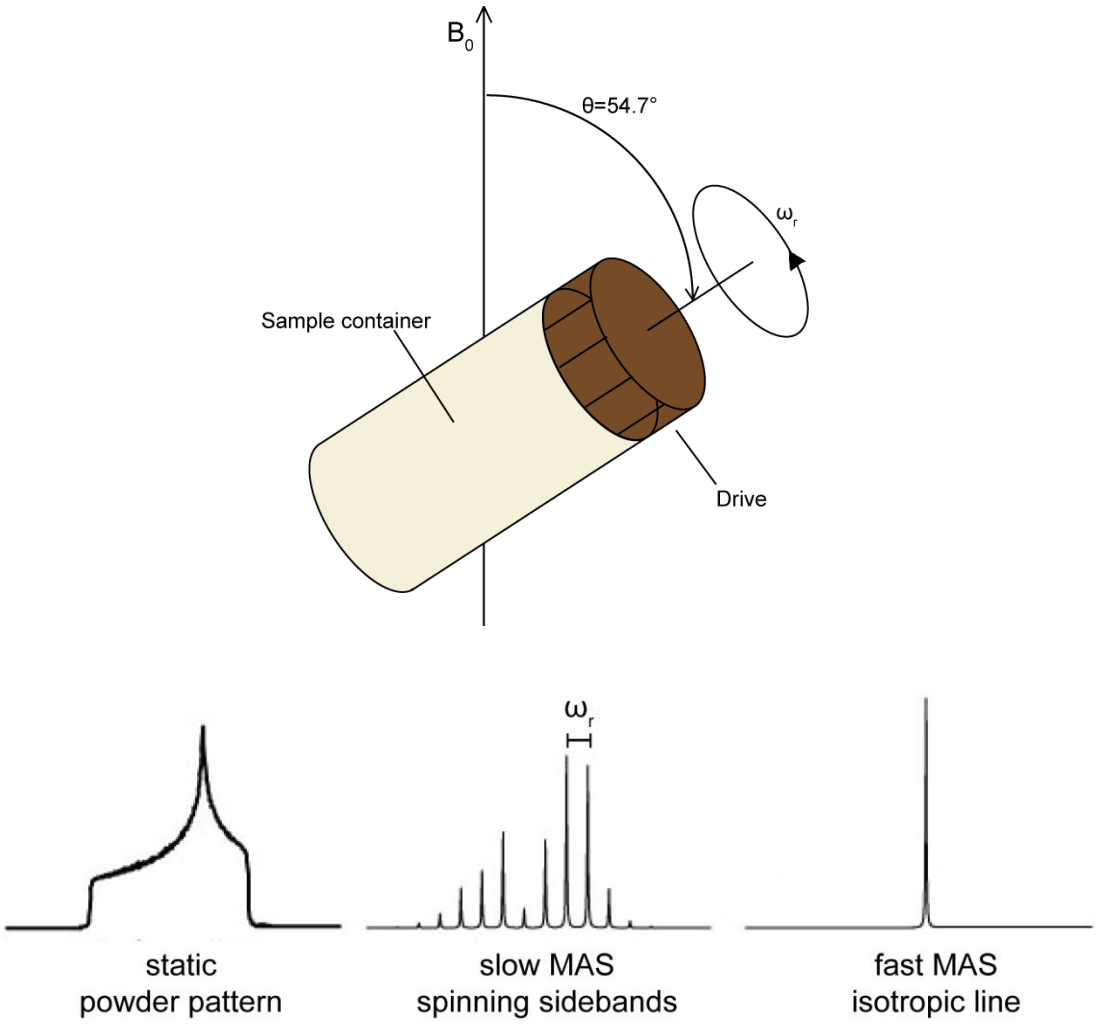


Figure 8: Magic angle spinning (MAS). a) MAS rotor: Samples are filled into ZrO cylinders with diameters between 1,3 to 7 mm and placed in the spectrometer with an angle θ of 54.7° to the magnetic field. Sample spinning is applied by an air stream towards the cogwheel at the rotor lid. b) MAS effect: static samples show a broad powder pattern due to anisotropic effects. At low spinning frequencies (ω_r) the spectrum splits up into a central line and several spinning sideband, separated from each other by ω_r. At fast MAS just the isotropic line and negligible spinning sidebands remain, while all anisotropic effects are averaged out.

Like exemplified in figure 8b, at slow MAS the powder pattern splits up into sharper resonances and spinning side bands while at faster MAS only the isotropic chemical shift is observed. Very high g-forces and heating act on the sample during MAS, so a balance between sufficiently high spinning rate and sample stability needs to be found. For all the experiments of this work, 3.2 mm MAS rotors and spinning frequencies between 10000 and 13500 Hz have been used. These MAS rates are not high enough to average out the strong dipolar couplings between protons, therefore just Carbon (^{13}C) and Nitrogen (^{15}N) detected spectra are feasible. For direct proton detection, ultra fast MAS (>60 kHz) has to be applied. Although MAS is extremely useful to improve the resolution in ssNMR experiments, many anisotropic interactions that contain important structural information like dipolar couplings are averaged out as well. Therefore pulse sequences that specifically reintroduce these information have to be applied (recoupling).

1.7.2 Cross polarization

The sensitivity of ^{13}C and ^{15}N detection is lower than ^{1}H detection due to the lower gyromagnetic ratio, lower natural abundance and longer T_1 relaxation times. A fundamental pulse sequence and starting point of most of the ssNMR experiments is cross polarization (CP). By magnetization transfer from protons to ^{13}C or ^{15}N the signal can be enhanced and, due to the faster relaxation time of protons, the repetition time can be abbreviated (Pines, Gibby et al. 1973). The scheme employs simultaneous irradiation of the protons and heteronuclei with matched radiofrequency fields. These “spin-lock” pulses result in a coherent precession of the spins. The magnetization is transferred if the energy difference of e.g. ^{1}H and ^{13}C spin states are equal, the so called Hartmann-Hahn matching condition (Hartmann and Hahn 1962). In practice, two irradiation pulses have to be optimized to match the energy difference on both nuclei. Usually a ramp shaped pulse is applied to one of the channels to improve signal stability and broaden the narrow match conditions. Cross polarization experiments are combined with high power dipolar decoupling of protons to increase the resolution. Several techniques of heteronuclear decoupling like TPPM (two phase pulse modulation) or SPINAL64 (small phase incremental alternation with 64 steps) are available (Bennett, Rienstra et al. 1995, Fung, Khitrin et al. 2000). The pulse sequence of a standard 1D CP-MAS experiment is shown in figure 9.

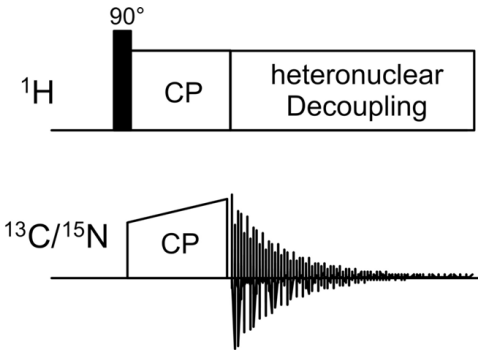


Figure 9: Pulse sequence 1D Cross Polarization. After a 90° pulse on protons, the magnetization is transferred to carbons by energy matched pulses on protons and carbon (cross polarization) using a ramped pulse on carbon. The signal is acquired on carbon and high power heteronuclear decoupling is applied to protons (e.g. TPPM, SPINAL64).

In this experiment, protons are first excited with a 90° pulse. Then, the magnetization is transferred by cross polarization using radio frequency pulses on both channels so that the Hartmann-Hahn condition is fulfilled. A ramped pulse is used on the low γ nucleus. Finally, the FID is detected under high power decoupling on protons. In comparison to a direct 90° pulse on ^{13}C or ^{15}N and consecutive detection, by CP, the signal is enhanced and the linewidth decreased.

1.7.3 Spin Diffusion

The standard two-dimensional experiment in solid state NMR is a ^{13}C-^{13}C correlation. The spectrum imparts information about the quality of the sample, not only about the signal-to-noise ratio but also about homogeneity, resolution and peak distribution. Furthermore it is used to identify isolated spin systems, spin system types and for the estimation of the number of ordered residues. After a 1D CP, the 2D ^{13}C-^{13}C correlation is usually the second spectrum to be measured on a new sample. Homonuclear correlation is achieved by exploiting the dense network of strongly dipolar coupled protons (Bloembergen 1949). During a mixing step in the pulse sequence, the so called Proton driven spin diffusion (PDSD) is mediated just by suspending the proton decoupling (Szeverenyi, Sullivan et al. 1982). If an additional rotor frequency synchronized pulse for recoupling of dipolar interactions is used during mixing, the method is called Dipolar assisted rotational resonance (DARR) (Takegoshi, Nakamura et al. 2001). The pulse sequence for a standard PDSD/DARR experiment is shown in figure 10.

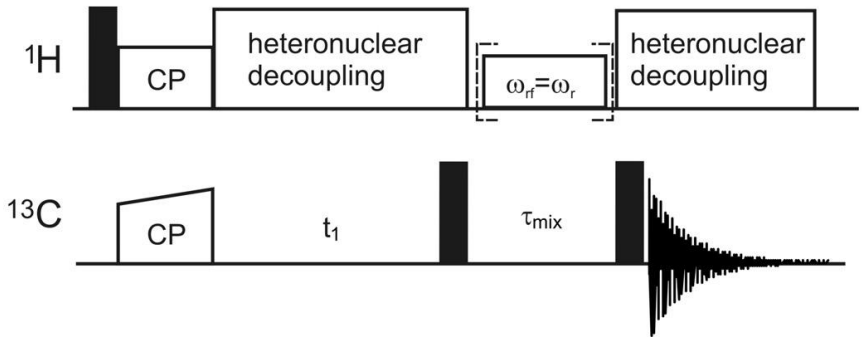


Figure 10: Proton driven spin diffusion pulse sequence for 2D ^{13}C-^{13}C correlation. After a ^{1}H-^{13}C CP and indirect evolution during t_1 a 90° pulse is used to store the magnetization along the z-axis. Then a spin diffusion step is applied for recoupling of ^{13}C-^{13}C dipolar interactions (τ_{mix}) and another 90° pulse is used to convert the carbon magnetization back to the transverse plane. Heteronuclear decoupling of protons is applied during t_1 and acquisition. In case of DARR, a weak recoupling field is applied during τ_{mix}. Filled bars represent 90° pulses. Mixing times range from 1-1000 ms.

After a 90° pulse on protons, the magnetization is transferred via CP to ^{13}C. The indirect carbon dimension is evolved during t_1, followed by the mixing step via PDSD or DARR ($\omega_{rf}=\omega_r$) and the acquisition time. High power heteronuclear decoupling is applied during t_1 evolution and acquisition. Essentially, all ^{13}C atoms within a certain distance to each other are correlated by cross peaks (beside the diagonal) in the spectrum. While with short mixing time (10-50 ms) all carbons within one spin system will be correlated, with longer mixing time (100-500ms) correlations between different residues will be detectable. Therefore, long mixing time ^{13}C-^{13}C

correlation spectra can be used to obtain sequential connectivity information between residues as well as for structural constraints (Castellani, van Rossum et al. 2002, Manolikas, Herrmann et al. 2008). An exemplified 2D DARR spectrum of HET-s (218-289) is shown in Figure 11. The mixing time of 50 ms results in intraresidue cross peaks correlating all side chain carbons to each other. Just rigid domains of the protein contribute to the spectrum. The correlated cross peaks of an isolated threonine and alanine residue are highlighted by dashed lines. For alanine, which has just one carbon in the side chain, one cross peak arises on both sides of the diagonal (Cα-Cβ) while for threonine, which contains two carbons in the side chain, the number of cross peaks on every side of the diagonal is increased to three (Cα-Cβ, Cα-Cγ, Cβ-Cγ).

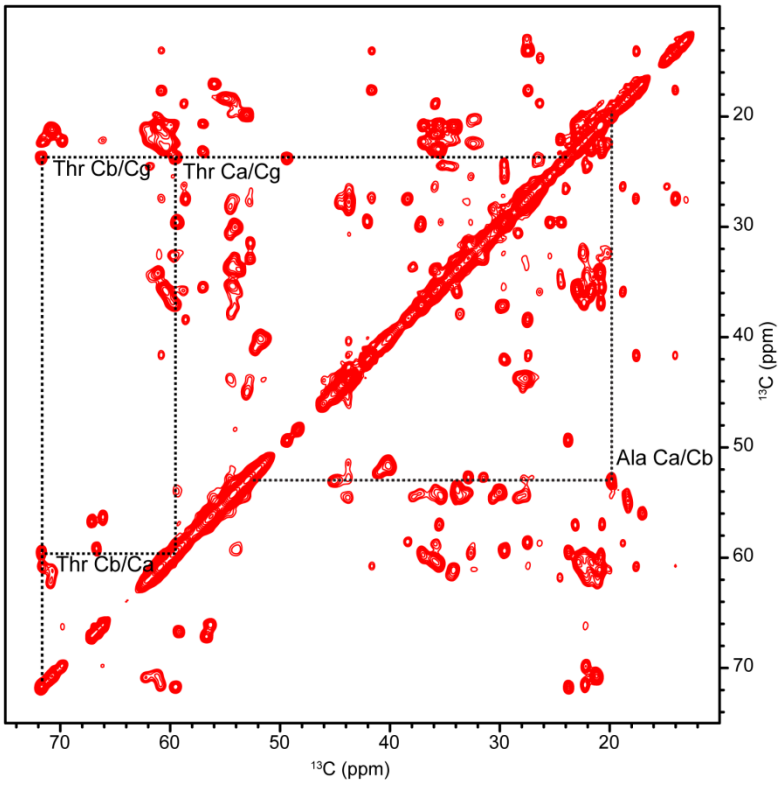


Figure 11: 2D DARR spectrum of HET-s(218-289) amyloid fibrils measured at 14.1T with a mixing time of 50 ms. Correlations of nuclei close in distance show up as cross peaks along the diagonal. Spin systems with a single carbon in the side chain (alanine) as well as two carbons (threonine) are exemplified.

1.7.4 Experiments for sequential assinments: Cross polarisation between low γ- nuclei

CP is not only used for signal enhancement, it is basically the common way to transfer magnetization between heteronuclei in ssNMR. For sequential assignments of protein backbone atoms, two and three dimensional ^{15}N-^{13}C correlation spectra are implemented which are based on a specific cross polarization. The magnetization is first transferred from protons to the low γ-nuclei nitrogen to enhance the nitrogen signal. To measure inter- or intra residue correlations the magnatization is the transferred either to the own Cα or to CO of the preceding residue. This transfer is feasible due to the different avarage chemical shift of Cα (~60 ppm) and CO (~175

ppm). For a specific CP from ^{15}N to either Cα or CO the ^{13}C frequency is set on resonance during CP (Baldus, Petkova et al. 1998). A rectangular pulse is applied on nitrogen, while on carbon usually either a ramped or adiabatic pulse is used. The chemical shift is evolved on the ^{15}N nuclei and detected on ^{13}C. As a result, a 2D NCα spectrum contains correlations of the backbone amide with the Cα of the same amino acid while a NCO spectrum correlates backbone amides with CO's of the preceding residue.

The basic spectra for the assigments of a protein backbone are NCACX and NCOCX (Pauli, Baldus et al. 2001). They are recorded using a combination of specific CP transfers and spin diffusion for ^{13}C-^{13}C correlation. To get the highest possible resolution, they are recorded as 3D spectra, but 2D versions without CA/CO evolution are also possible. The pulse sequence of a 3D NCACX is shown in figure 12. The magnetization is transferred from proton to nitrogen via cross polarization and then, selectively, either to CA or CO. A PDSD or DARR step is then used to transfer magnetization to other surrounding carbons.

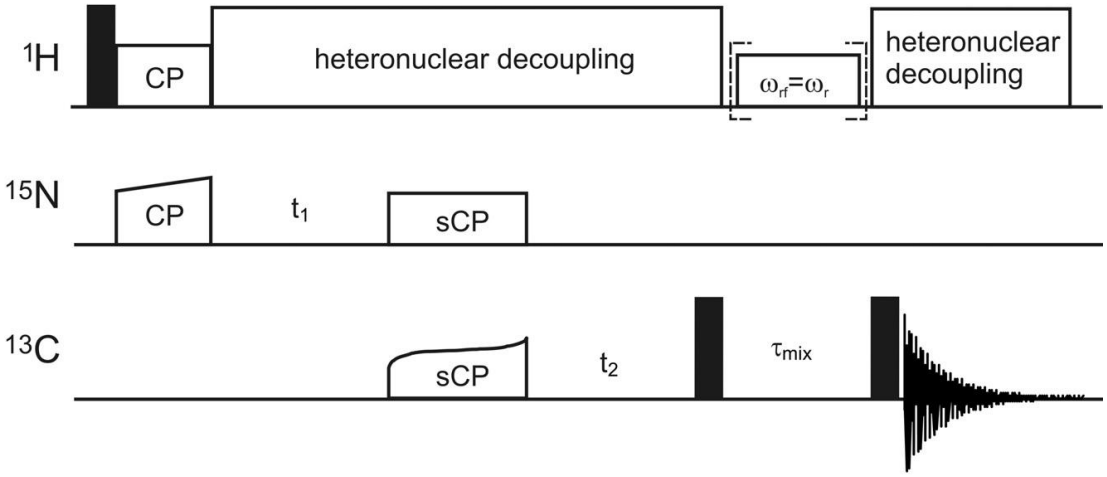


Figure 12: Pulse sequence of a 3D ^{15}N-^{13}C correlation. After ^{1}H-^{15}N CP and nitrogen evolution t_1, the magnetization is transferred by a specific CP to either CA or CO which is evolved during t_2. A PDSD or DARR mixing is used for ^{13}C-^{13}C correlation. High power heteronuclear decoupling of protons is applied during t_1, t_2, the specific CP and acquisition. Filled Bars represent 90° pulses.

1.7.5 Sequential sequence assignments

A sequence specific resonance assignment is the first step towards structural information of a protein. It requires homogeneous samples and the measurement of well resolved spectra. The minimum set of experiments includes a 2D ^{13}C-^{13}C correlation spectrum as well as NCOCX and NCACX spectra (2D or 3D). In a NCACX with a short mixing time, intra-residue correlations are detected, for example N-CA-CB and N-NA-CG of one residue. In a NCOCX with short mixing time, correlations from a backbone amide to the preceding CO and then to CA and CB of the same preceding residue are detected. The magnetization transfers of both experiments are illustrated in figure 13. The NCOCX provides complementary sequential information to NCACX. The assignment procedure initiates spin system identification using 2D ^{13}C-^{13}C spectra and NCACX spectra. As soon as at least the N, CA, and CO of a number of spin systems are assigned, the sequential connections can be found by matching the assigned carbon shifts to the NCOCX spectra. This dataset is the minimal requirement for sequence specific assignments

and sufficient for small proteins and peptides. A useful addition is a CANCO experiment providing quasi through bond sequential information (Li, Berthold et al. 2007). Several pulse sequences for assignments of larger molecules have been developed as well so that nowadays proteins with several hundreds of residues can be readily assigned (Schuetz, Wasmer et al. 2010). Due to spectral crowding and overlap the assignment of uniformly labeled samples is challenging. Most of the times a combinatory approach of multidimensional NMR and isotope labeling techniques is utilized (Renault, Cukkemane et al. 2010). Advantages and disadvantages of different isotope labeling strategies will be discussed in chapter 1.9.

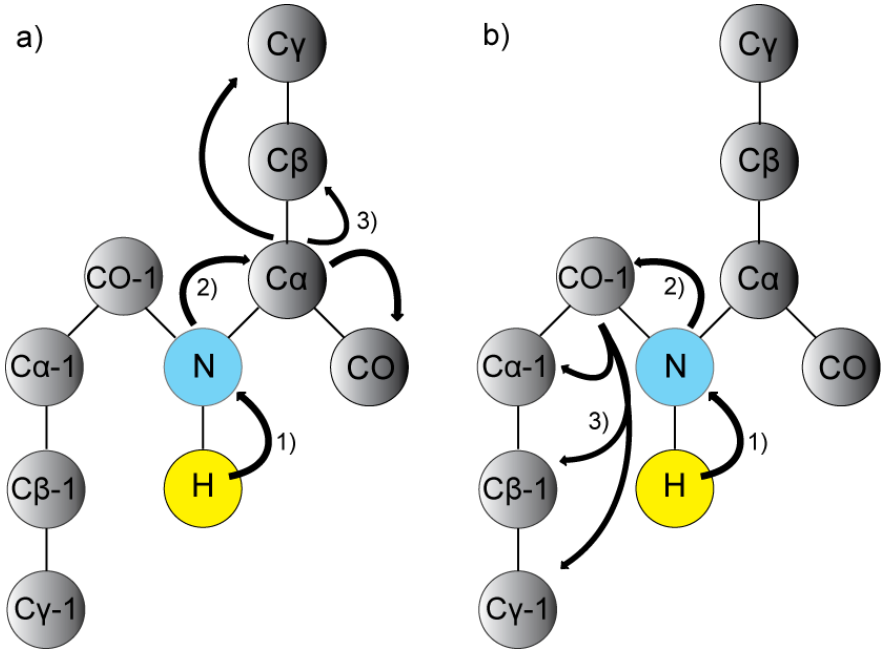


Figure 13: Polarization transfers in NCACX, NCOCX experiments. a) Magnetization is transferred from ^{1}H to amide nitrogen and further to CA. By spin diffusion all carbons within the same residue are correlated. b) Magnetization is transferred from ^{1}H to amide nitrogen and further to CO of the previous residue (CO-1). By spin diffusion all carbons within the previous residue are correlated. With long mixing times during spin diffusion sequential connections can be detected.

1.7.6 Secondary structure from Chemical shifts

Residue specific secondary structure information can be obtained from protein backbone chemical shifts. Especially the CA and CB chemical shifts changes are remarkable, in comparison to random coil chemical shits, if the residue is involved in a secondary structure element. In case of a α-helix the Cα chemical shift is usually increased while the Cβ chemical shift of the same residue is decreased. For a β-sheet it's the opposite situation. The secondary structure specific chemical shift deviations of Cα and Cβ are illustrated in figure 14. Software like TALOS can use the chemical shifts to make an empirical prediction of the backbone torsion angles which can be used in the structure calculation (Cornilescu, Delaglio et al. 1999).

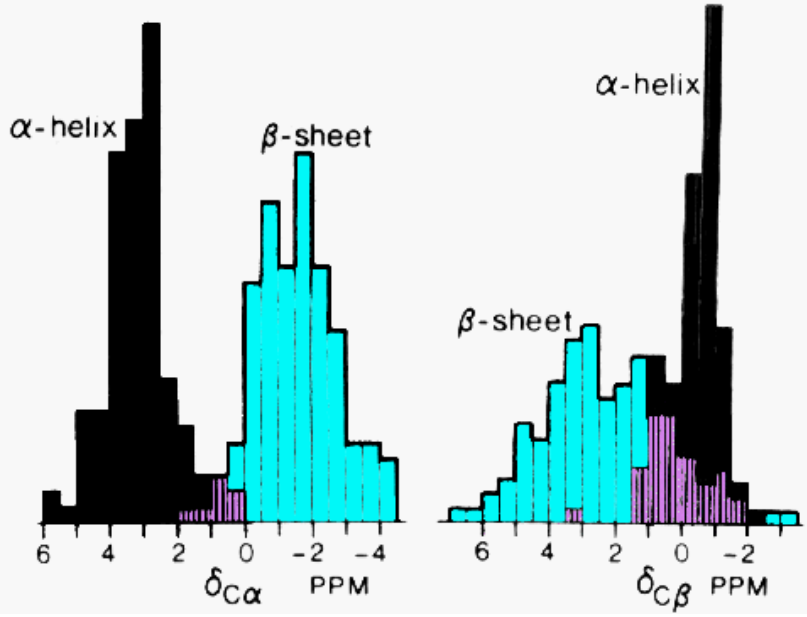


Figure 14: Chemical shift deviations of backbone nuclei depending on secondary structure elements, compared to random coil shifts in ppm (reference: http://spin.niddk.nih.gov/bax/software/TALOSORIG/).

1.8 Protein structure calculation from NMR data

There are three main steps that have to be followed towards a NMR structure: a) sequential resonance assignments, like described earlier, b) collection of structural restraints and c) the structure calculation itself. During the assignment process each residue nuclei is attributed with a chemical shift value. The correct assignment is the critical step in the progress and the chemical shift data can already be used for structure modeling using software such as CS-ROSETTA (Shen, Lange et al. 2008). Several types of experiments can be utilized and should be combined to obtain distance restraints: Experiments that use long spin diffusion steps such as 2D PDSD or 3D NCACX spectra with mixing times of up to 500 ms (Castellani, van Rossum et al. 2002, Castellani, van Rossum et al. 2003), experiments that indirectly detect ^{1}H-^{1}H distances such as CHHC and NHHC (Lange, Luca et al. 2002) or experiments that are based on third spin assisted recoupling such as PAR and PAIN (De Paepe, Lewandowski et al. 2008, De Paepe, Lewandowski et al. 2011). Other types of restraints that can be added to the structure calculation are e.g. dihedral angles, hydrogen bonds and disulphide bridges.

The structure calculation is finally aiming to fulfill the detected constrains without violating chemical properties such as force fields, bond length and angles. Molecular dynamics algorithms such as CNS, X-PLOR or DYANA are used to find a global minimum of a hybrid energy function (Guntert, Mumenthaler et al. 1997, Brunger, Adams et al. 1998, Schwieters, Kuszewski et al. 2003). Starting from an extended polypeptide chain the potential energy is minimized and an ensemble of 10-20 structures with the lowest energy is usually used to represent the NMR structure. In the end an iterative refinement and validation of the structure needs to be performed.

1.9 ISOTOPE LABELING STRATEGIES

Due to the marginal natural abundance of ^{13}C and ^{15}N, for NMR spectroscopic investigations proteins have to be enriched with isotopes (reviewed in (Renault, Cukkemane et al. 2010)). There are several ways to achieve this, but here just the most common, using recombinant protein expression in *E.coli* will be highlighted. The method relies on feeding bacteria with specific isotope labeled nutrients. A simple minimal media usually contains a single carbon and nitrogen source, magnesium sulfate, vitamins, trace metals and Phosphates as buffer substance. While as nitrogen sources ^{15}N labeled ammonium salts are used, for ^{13}C labeling glucose or its derivatives (glycerol, acetate, pyruvate, succinate) are used as sole carbon source. Nitrogen labeling is the cheapest and simplest form of protein labeling. It is very common for initial experiments of solution NMR to verify the protein folding and quality of the preparation with 2D ^{1}H-^{15}N spectra. For solid state NMR the most common scheme is uniform ^{13}C-^{15}N labeling by the use of fully labeled glucose and ammoniumchloride (McIntosh and Dahlquist 1990). Uniform labeling generates the highest amount of spectroscopically detectable information. However, with larger proteins or repetitive primary sequence, the spectra may suffer from spectral overlap.

The so called “reverse” labeling, in which unlabeled amino acids are added to the minimal medium, can eliminate amino acid types from the spectra and therefore facilitate the assignment of the others. In contrast, “forward” labeling can be applied by the use of a defined minimal medium containing labeled amino acids. This is particularly useful if only specific amino acid types need to be investigated. A characteristic distribution of ^{13}C and ^{12}C isotopes can be achieved by the use of position specific isotope labeled carbon sources like [2-^{13}C]-glycerol, [1,3-^{13}C]-glycerol, [1-^{13}C]-glucose or [2-^{13}C]-glucose. This so called checkerboard labeling not only reduces spectral crowding, but also eliminates most of the J-couplings which results in sharper lines. Additionally, the strong dipolar couplings between neighboring nuclei are removed; by that, the measurement of weak long range distance restraints is facilitated. For amino acids synthesized by the citric acid cycle, several isotopomers are obtained and the exact degree of isotope labeling at these positions needs to be analyzed for every new sample (LeMaster and Kushlan 1996, Castellani, van Rossum et al. 2002, Higman, Flinders et al. 2009, Loquet, Giller et al. 2010). By segmental isotope labeling, just a certain part of a protein is labeled with isotopes. This is realized by protein trans-splicing or chemical ligation of peptides. Spectral overlap can be significantly reduced and the assignment of large proteins or repetitive sequences permitted. The structure and function of particular domains of a protein can be exploited (Muralidharan and Muir 2006). The methodology is not entirely straight forward and will be discussed to more detail in the next chapter.

In certain cases it is complicated to differentiate between inter- and intramolecular distance restraints. This is an issue especially for continuous protein arrays like amyloid fibrils, but is a challenge also for regularly assembled protein complexes in general. Inter- and intramolecular restraints should be unambiguously assigned. Entirely intramolecular restraints for a structure calculation of HET-s(218-289) have been obtained by sample dilution with unlabeled material

(Wasmer, Lange et al. 2008). Intermolecular distances can be measured on “mixed” samples. The simplest way is using 50% ^{15}N + 50% ^{13}C samples and measurements of N-C correlation spectra. Recently, also mixed [1-^{13}C]-glucose and [2-^{13}C]-glucose labeled samples were used to obtain unambiguous intermolecular contacts (Loquet, Giller et al. 2010). Another option would be the use of segmental labeling by labeling only one part of the protein to detect exclusively intramolecular distances or to label one part with ^{13}C and another with ^{15}N and collect interdomain distance restraints.

1.10 SEGMENTAL ISOTOPE LABELING OF PROTEINS

Segmental isotope labeling relies on the ligation of an isotope labeled and an unlabeled polypeptide of the same protein. For very small proteins it can be achieved by using fully synthetic peptides and native chemical ligation (NCL) (Kent 2009).

Isotope labeling strategies that involve bacterial protein expression are more sophisticated, such as the case of expressed protein ligation (EPL) which is a combination of recombinant expression and synthesis (Muir, Sondhi et al. 1998). This approach is very useful to introduce chemical modifications at specific positions like fluorophores or posttranslational modifications. However, NCL and EPL are limited by the size of the synthetic peptide and additionally dependent on the introduction of a α-thioester which is challenging. The most suitable ligation technique for segmental isotope labeling of proteins is protein splicing, an autocatalytic process based on an intervening protein domain (intein) that excises itself from the polypeptide in which it is embedded, concomitantly creating a new peptide bond between its two flanking regions (exteins) (Vila-Perello and Muir 2010). Some inteins have been found to be naturally split, in this case the reaction is named protein trans-splicing. Their primary sequence is cut into two polypeptides giving an N-terminal fragment (IntN) and a C-terminal fragment (IntC). The complementation of the intein domains leads to the reconstitution of the canonical fold and the recovery of splicing activity (Wu, Hu et al. 1998). Beside the natural ones, also several artificially split inteins have been generated, generally in order to obtain smaller intein Fragments that provide access to chemical synthesis.

The first discovered and best described naturally split intein is the DnaE intein from *Synechocystis sp*., but the intein with the highest activity under various conditions and the highest tolerance to flanking amino acids is the DnaE intein from *Nostoc punctiforme* (Wu, Hu et al. 1998, Iwai, Zuger et al. 2006). All inteins originate from the same ancestral gene and share a highly conserved three dimensional structure. The same fold is observed in hedgehog proteins and is therefore called HINT-fold (Hedgehog-Intein) (Hall, Porter et al. 1997). The solution structure of the *Npu* DnaE intein has been solved as a single chain variant which is illustrated in figure 15, showing the IntN domain in blue and the IntC domain in red (Oeemig, Aranko et al. 2009). It consists mainly of β-strands with two short helices in the IntN fragment. The complex shares intermolecular β-sheets at several positions and adopts a fold so that the splicing junction (N+C) allies close in space. Most commonly, catalytic cysteine residues are located at

the splicing junction but inteins with catalytic serine or threonine have also been described (Telenti, Southworth et al. 1997).

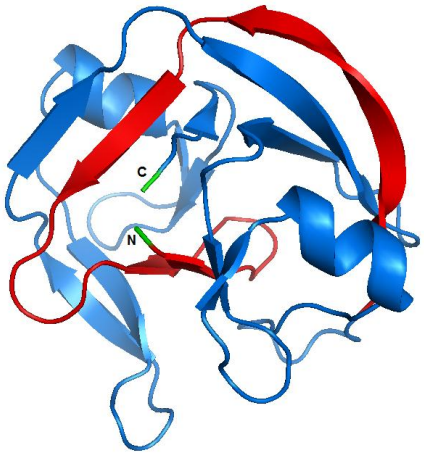


Figure 15: Solution structure of the *Npu* DnaE intein complex (derived from PDB ID: 2KEQ). The 102 residue IntN domain is shown in blue, the 32 residue IntC domain is shown in red.

The protein trans-splicing reaction occurs by the following mechanism, a model is shown in figure 16. After the intein fragment association and folding into the active state, the scissile bond at the N-terminal splicing junction is activated by an N-S acyl shift to a linear thioester. This intermediate is attacked by the nucleophilic thiol group of the C-extein to form a branched intermediate. Succinimid formation of asparagine at the splicing junction cleaves the peptide bond between the intein complex and the C-extein. The native peptide bond is reassembled by an S-N acyl shift of the cysteine residue (Xu, Southworth et al. 1993). Potential side effects are the cleavage of intein fragments by a reaction of intermediate states with a nucleophilic buffer component. For segmental isotope labeling, intein fusion proteins of individual domains of a protein of interest have to be generated. Those can either be expressed and purified separately while only one part is labeled with isotopes and subsequently spliced *in vitro* or be consecutively expressed in the same cell under control of different promoters to achieve a splicing reaction *in vivo* (Muona, Aranko et al. 2010). The strategy is proved for several soluble proteins and led to the almost complete backbone assignment of the 52 kDa F_1-ATPase (Yamazaki, Otomo et al. 1998, Yagi, Tsujimoto et al. 2004). The approach has even been expanded with an engineered intein to label a central segment of a protein (Busche, Aranko et al. 2009). Although the demand for elaborated labeling strategies in solid state NMR is high due to generally broad lines and tendency to overlap, segmental isotope labeling of insoluble proteins has not yet been demonstrated.

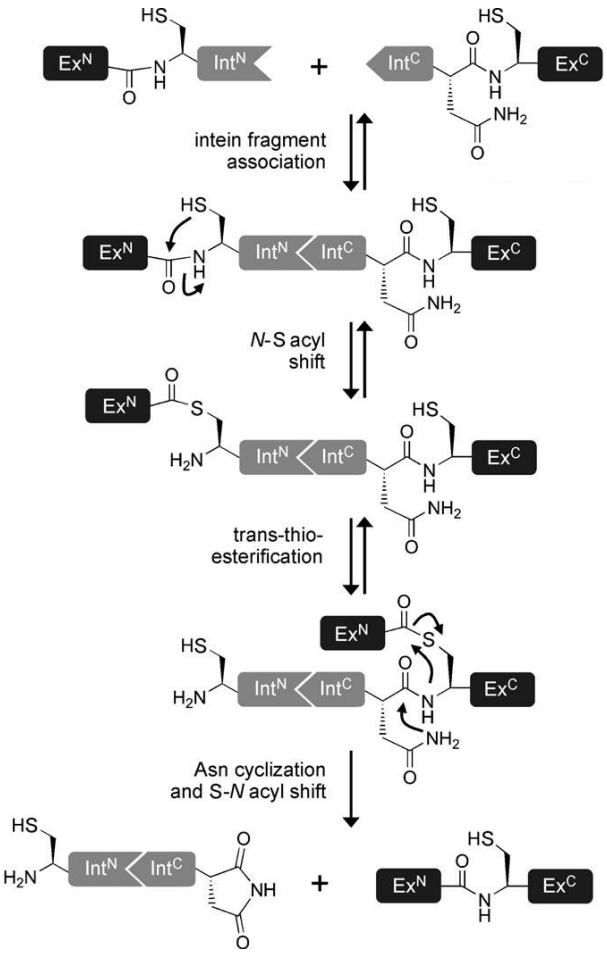


Figure 16: Mechanism of protein trans-splicing (reproduced from (Mootz 2009) with permission from Wiley). The intein is split into two parts, IntN and IntC and fused with extein protein domains, ExN and ExC respectively. After the intein fragments associate and fold into the active state, the splicing reaction is initialized by a N-S acyl shift to form a thioester intermediate. This is attacked by the thiol group of the cysteine residue of ExC and gives rise to a branched intermediate. By cyclization of the asparagine residue at the splicing junction the peptide bond between IntC and ExC is cleaved. During the succinimid formation of the asparagine a S-N acyl shift occurs and results in a native peptide bond between ExN and ExC.

2 AIMS AND SCOPE

Enterobacterial curli are functional amyloids involved in biofilm formation and host cell binding. Their biogenesis is well characterized and the amyloid formation was studied *in vivo* as well as *in vitro*. The major curli subunit CsgA forms amyloids that share all characteristic features with disease related amyloids and is therefore a potential system to gain deeper insight to amyloid structure, functions and mechanisms. Residue specific structural information have already been obtained by quenched H/D exchange analysis, but high resolution data was not available.

The aim of this thesis was the structural characterization of CsgA fibrils by solid state NMR. The requirements for structure elucidation by solid state NMR are an almost complete backbone and side chain ^{13}C resonance assignment as well as the detection of long range distance restraints. The feasibility of a complete ssNMR analysis relies on samples with high signal to noise ratio and good resolution. Therefore highly pure and homogeneous ^{13}C-^{15}N labeled CsgA amyloid fibrils had to be produced and high resolution NMR data to be measured and analyzed.

Furthermore segmental isotope labeling by protein *trans* splicing was applied and had to be adjusted for insoluble proteins to facilitate the assignment of highly overlapped spectra and the detection of unambiguously intramolecular distance restraints. The strategy was demonstrated for amyloid fibrils of the major curli subunit CsgA as well as the well characterized HET-s (218-289). Therefore, appropriate intein fusion constructs had to be designed; expression and purification protocols to be established. Moreover the conditions for the splicing reaction needed to be optimized and the strategy for the product purification had to be improved.

The labeling strategy enabled a partial assignment of CsgA by removing highly overlapped peaks from the spectra. Unambiguous distance restraints were detected both for CsgA and HET-s.

3 MATERIALS AND METHODS

3.1 STANDARD MATERIALS

3.1.1 CHEMICALS, ENZYMES, ANTIBODIES AND KITS

If not stated otherwise, all chemicals that were used in this work were of *"pro analysis"* grade. Chemicals were purchased from the companies Amersham Biosciences, Fluka, Merck, Millipore, Omnilab, Qiagen, Roche, Roth, Sigma and Stratagene.
Enzymes (Table 4), antibodies (Table 5 and Table 6) and kits (Table 7) were used as described by the corresponding company.

Table 4: Enzymes used in this work

Enzyme	Source	Specification
Hen white-egg lysozyme	Fluka	Glycoside hydrolase
DNAse	Merck	Endodeoxyribonuclease
Platinum *Pfx* Polymerase	Invitrogen	DNA polymerase
DreamTaq Polymerase	Fermentas	DNA Polymerase
T4 DNA Ligase	New England Biolabs	DNA Ligase
BamHI, HindIII, NheI, NdeI, NcoI, XbaI, XhoI	New England Biolabs, Fermentas	Restriction endonucleases

Table 5: Primary Antibodies.

Name	AB species	Antigen	Dilution	Source
His	Mouse	Hexa-His	1:1500	Novagen

Table 6: Secondary Antibodies.

Name	AB species	A/R species	Conjugate	Dilution	Source
Anti Mouse IgG (H+L) AP Conjugate	Goat	Mouse	Alkaline phosphatase	1:7500	Promega
Anti Mouse IgG (H+L) HRP Conjugate	Goat	Mouse	Horse radish peroxidase	1:2000	Dianova

Table 7: Kits used in this work

Kit	Source	Usage
QIAprep Spin Miniprep Kit	Qiagen	Plasmid purification
QIAquick PCR Purification Kit	Qiagen	Purification of PCR products, removal of nucleotides
QIAquick Gel Extraction Kit	Qiagen	DNA extraction from gel slices
Lumi-Light Western Blotting Substrate	Roche	Detection of HRP conjugated antibodies by ECL
BCIP/NBT Color Development Substrate	Promega	Detection of AP conjugated antibodies

3.1.2 Molecular weight standards for electrophoresis

Following molecular weight standards were used during this work.

Table 8: Molecular weight standards.

Name	Source	Type
Smart Ladder	Eurogentec	Agarose gel DNA marker
Unstained Protein Molecular weight marker	Fermentas	SDS-PAGE protein marker
PageRulerTM Unstained Protein ladder	Fermentas	SDS-PAGE protein marker
PageRulerTM Prestained Protein ladder	Fermentas	SDS-PAGE protein marker

3.1.3 Bacterial Strains

Table 9: Bacterial strains used for molecular cloning or recombinant protein expression.

E. coli strain	Genotype	Source
BL21 (DE3)	*F– ompT gal dcm lon hsdSB(rB- mB-) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])*	Stratagene
DL39	*F– Lam–acpC13 fnr-25rph-1 ilve12 tyrB507*	LeMaster and Richards, 1988
DH5	*F– endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK- mK+), λ–*	Invitrogen
T7 Express	*fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--*TetS*)2 [dcm] R(zgb-210::Tn10--*TetS*) endA1 Δ(mcrC-mrr)114::IS10*	NEB

3.1.4 Plasmids

Following Plasmids were used during this work. Those constructs generated in this work were cloned into pET11a, pET28c, pET21, pET44a, pGEX6c or pET20M, all obtained from Joop van den Heuvel.

Table 10: Initial and generated plasmids.

Plasmid	Features	Source
pET-11d CsgA21-151	pET-11d derivative encoding *E. coli* CsgA21-151	A. Eberth; HZI, Braunschweig
pET-21-HET-s218-289	pET-21 derivate encoding *P. anserine* HET-s218-289	C. Ritter; HZI, Braunschweig
pAU08(gpD-IntN) (Zettler et al. 2009)	pETDuet derivate encoding gpD- *Npu* $DnaE^{N}$ (IntN)	H. Mootz, Münster
pVS01(IntC-Trx) (Zettler et al. 2009)	pBAD derivate encoding *Npu* $DnaE^{C}$ (IntC)-Trx	H. Mootz, Münster
pET-28c HETs-IntN	pET-28c derivate encoding HET-s218-256-IntN	This work
pET-11d IntC-HETs	pET-11d derivate encoding IntC-HET-s257-289	This work
pET-11d His-CsgAR1-IntN	pET-11d derivate encoding CsgA21-70-IntN	This work
pET-11d His-CsgAR12-IntN	pET-11d derivate encoding CsgA21-90-IntN	This work

pET-11d His-CsgAR1234-IntN	pET-11d derivate encoding CsgA21-132-IntN	This work
pET-28c CsgAR1-IntN-His	pET-28c derivate encoding CsgA21-70-IntN	This work
pET-28c CsgAR12-IntN-His	pET-28c derivate encoding CsgA21-90-IntN	This work
pET-28c CsgAR1234-IntN-His	pET-28c derivate encoding CsgA21-132-IntN	This work
pET-11d His-IntC-CsgaR2345	pET-11d derivate encoding IntC-CsgA71-152	This work
pET-11d His-IntC-CsgaR345	pET-11d derivate encoding IntC-CsgA91-152	This work
pET-11d His-IntC-CsgaR5	pET-11d derivate encoding IntC-CsgA133-152	This work
pETM20 Trx-IntC-R5	pETM20 derivate encoding Trx-IntC-CsgA133-152	This work
pET44 Nus-IntC-R5	pET-44 derivate encoding Nus-IntC-CsgA133-152	This work
pGEX6 GST-IntC-R5	pGEX6c derivate encoding GST-IntC-CsgA133-152	This work

3.1.5 Oligonucleotides

Oligonucleotides were used as primers to amplify genes of interest by Polymerase chain reaction for molecular cloning. The primers were purchased from Eurofins MWG.

Table 11: Oligonucleotides used for molecular cloning

Name	Sequence 5‘-3‘	Restriction Site
Nde-6H-CsgA for	ATATCATATGCATCACCATCACCATCAC GGTGGTTGTTCCTCAGTACG	NdeI
CsgA T70 rev	CAAGTCAGAGTTACGGGCA (pho)	-
CsgA S90 rev	GCTGTCATCTGAGCCCTG (pho)	-
CsgA S131 rev	TGCAGTCTGGTCAACTGCA (pho)	-
IntN for	TGTTTAAGCTATGAAACGGAAATATTG (pho)	-
IntN - BamHI rev	TATAGGATCCTTAATTCGGCAAATTATCAACCCG	BamHI
Nde-6H-IntC for	TATACATATGCATCACCATCACCATCACATC AAAATAGCCACACGTAAATATTT	NdeI
IntC-rev	ACAATTAGAAGCTATGAAGCCATTTTTG	-
CsgA 71 for	ATTACCCAGCATGGCGGCG	-
CsgA 91 for	ATCGATCTGACCCAACGTGG	-
CsgA 132 for	AACTCCTCCGTCAACGTGACT	-
CsgA BamHI rev	TATAGGATCCTTAGTACTGATGAGCGGTCGCG	BamHI
bam tev intcr5 for	TATAGGATCCGAAAACCTGTATTTTCAGGGC ATCAAAATAGCCACACGTAAATATTTAG	BamHI
intcr5-rev	TATAAAGCTTTTAGTACTGATGAGCGGTCGCGTTGTTA	HindIII
Nco CsgA for	TATACCATGGGTGTTGTTCCTCAGTACGGCG	NcoI
IntN Xho rev	TATACTCGAGATTCGGCAAATTATCAACCCGCATCA	XhoI
Nde-hets-for	ATATCATATGAAGATCGACGCGATTGTGGGAAGG	NdeI
IntN Xho rev	TATACTCGAGTTAATTCGGCAAATTATCAACCCG	XhoI
Hets nterm rev	GATACGGATCCCACCATGTAG	-
pet 11 IntC for	TTAAGGATCCGGCTGCTAACAAAGCCCG	BamHI
D258 cterm for	GACCAGACAACCAACTCGGTA	-
cterm rev	TATAGGATCCTTAATTATCCCAGAACCCCTTACCT	-

3.1.6 Media and Buffer

The media and electrophoresis buffers, which were used in this study, are summarized in Table 12 and Table 13. All media were sterilized by autoclaving (121 °C, 2 bar, 20 min, Top7000PST, Sauter) and heat-sensitive additives were sterile filtered (pore width 0.2 µM). Antibiotics were added after media cooling below 50 °C. Depending on plasmid and bacterial strain the following antibiotics were used: ampicillin (100 µg/ml) or kanamycin (20µg/ml). All other buffers are mentioned in their associated passage.

Table 12: Bacterial growth media used in this study.

Medium	Composition
Lysogeny Broth (LB)	1 % (w/v) tryptone, 0.7 % (w/v) NaCl, 0.5 % (w/v) yeast extract, adjusted to pH 7.5 with NaOH
LB-Agar	LB, 1.5 % (w/v) agar-agar
Terrific Broth (TB)	1.2 % (w/v) tryptone, 2.4 % (w/v) yeast extract, 0.4 % (v/v) glycerol, 0.1 x TB salts
TB salts	0.17 M KH_2PO_4, 0.72 M K_2HPO_4
Minimal Medium (CN040) (1L)	50 mM K_2HPO_4, 50 mM KH_2PO_4, 1g NH_4Cl, 4g glucose, 2 mM $MgSO_4$, 5mM Na_2SO_4, 200µl trace metal mix, 10ml MEM vitamin solution (SigmaAldrich)
Trace metal mix	100 µM $FeCl_3$, 40 µM $CaCl_2$, 20 µM $MnSO_4$, 20 µM $ZnSO_4$, 4 µM $CoCl_2$, 4 µM $CuCl_2$, 4 µM $NiCl_2$, 4 µM Na_2MoO_4, 4 µM H_3BO_3

Table 13: Buffers used for Agarose Gel electrophoresis, SDS-PAGE and Western Blot.

Solution/Buffer	Composition
AP buffer	100 mM Tris/HCl pH 9.5, 100 mM NaCl, 5 mM MgCl2
Blocking solution	TBST, 5 % (w/v) skimmed milk powder
Coomassie staining solution	30 % (v/v) ethanol, 10 % (v/v) acetic acid, 0.25 % (w/v) Coomassie R-250
Coomassie destaining solution	40 % (v/v) ethanol, 10 % (v/v) acetic acid
6 x DNA loading buffer	10 mM Tris/HCl pH 7.4, 25 mM EDTA, 30 % (v/v) glycerol,
1 x TAE buffer	40 mM Tris, 20 mM sodium acetate, 1 mM EDTA, adjusted to pH 8.2 with acetic acid
1 x TBST	20 mM Tris/HCl pH 8, 150 mM NaCl, 0.05 % (v/v) Tween-20
Transfer buffer for Western Blotting	20 mM Tris pH 8, 192 mM glycine, 15 % (v/v) Methanol
Y x SDS sample buffer	0.1 M Tris/HCl pH 6.8, 2 % (w/v) SDS, 0.2 % (w/v) Bromphenol blue, 20 % (v/v) glycerol, 5 mM DTT
1 x NuPAGE SDS running buffer	50 mM MES, 50 mM TRIS, 3.5 mM SDS, 1mM EFTA pH7.3

3.2 Molecular Biology Methods

Molecular biology methods, which were applied in this study, were adapted from standard procedures (Sambrook and Russell 2000, Coligan 2003, Ausubel, Brent et al. 2007) and are not described in detail unless essential modifications were made. More detailed descriptions of biophysical methods are available from textbooks (Serdyuk, Zaccai et al. 2007).

3.2.1 TRANSFORMATION OF PLASMID DNA

In order to cultivate bacteria for recombinant protein expression or plasmid augmentation for molecular cloning, chemically competent *E.coli* cells were transformed with plasmid DNA via heat shock and subsequent cultivation on agar plates. Between 40-100 ng Plasmid DNA were used for 50-100 µl bacterial cell suspension. An antibiotic resistance gene located on the plasmid was used for positive selection of the transformation. One hour cultivation in antibiotic free liquid medium is necessary prior to plating in case of kanamycin resistance. For liquid cultures an appropriate volume of culture medium was inoculated with a single colony. These cultures were incubated at 37° in baffled flasks under continuous transverse agitation at 130 rpm. Plasmids were extracted from DH5α or XL1-Blue cells. *E.coli* BL21(DE3), DL39 or T7 Express cells were cultured for recombinant gene expression.

3.2.2 ISOLATION OF PLASMID DNA

The isolation of plasmid DNA is carried out by alkaline lysis of *E. coli* cells described by Birnboim and Doly (Birnboim and Doly, 1979) and distributed as the Miniprep Kid (Quiagen, Hilden). The bacterial cells get lysed with SDS, and the DNA denatured with sodium hydroxide. The lysate is neutralized by adding potassium acetate which leads to precipitation of denatured proteins and chromosomal DNA with the potassium salt of dodecylsulfate. The low molecular weight plasmids remain in solution and renature in the neutral conditions. The insoluble components are separated by centrifugation and the plasmid DNA can be isolated from the supernatant. For this purpose, a DEAE anion exchange chromatography column is used which binds the negatively charged DNA. After washing, the pure Plasmid DNA is eluted in pure water or low concentrated TRIS buffer.

3.2.3 DNA ANALYSIS

Horizontal agarose gel electrophoresis is used for the analysis and purification of DNA fragments. Agarose is a polysaccharide from red seaweed which can get dissolved in water by boiling and forms gels at decreased temperature. Relatively large pores remain in the gels which are exploited to separate DNA molecules according to their molecular weight. Samples, diluted with 6x sample buffer, are separated in 0.8% agarose gels by applying an electric field corresponding to 5 V/cm for 60 min in TAE buffer. Staining of the sample bands is carried out using the intercalator ethidium bromide (4µl in 50ml gel) which can be documented under UV illumination at 254 nm. For purification, DNA bands of desired size are excised from the gel, and worked up by melting the gel fragment and purify the DNA as described for before.

3.2.4 Polymerase chain reaction (PCR)

PCR is a technique for exponential amplification of genes of interest by cyclic binding, prolongation and denaturation of short primers on a template DNA. The amplification is accomplished using a polymerase that is active at high temperatures and automated cycling from ~60°C for annealing, 72°C for amplification and 90°C for denaturation. The annealing temperature is dependent of the melting temperature of the primers. The Platinum Pfx Polymerase (Invitrogen) in 50 µl sample volume was usually used for cloning and DreamTaq Polymerase (Thermo Scientifc) in 10 µl samples for colony PCR screens.

3.2.5 Restriction digestion of DNA

The restriction, the "cutting" of the DNA double strands is performed using restriction enzymes by incubating at 37°C. Several restriction enzymes cut DNA double strands in specific sequences which are usually palindromic. This can result in blunt ends or "sticky ends". The amount of enzyme is dependent on the amount of DNA and the buffer is provided by Manufacturer. After digestion the fragments are purified over an agarose gel.

3.2.6 Ligation of DNA

The ligation of DNA fragments is obtained by use of T4 DNA ligase. In case of sticky end ligation, the DNA strands need to be digested with the same restriction enzymes, so that the terminal bases are compatible to each other. The most regular case is the insertion of a PCR product into a expression vector. Generally 25 ng vector is used and the amount (ng) of required PCR fragment is calculated after 125 x Bp (PCR) / Bp (Vector).
PCR products amplified by phosphorylated primers can be blunt end ligated by the use T4 DNA Ligase. In comparison to sticky end ligation it is less specific and can result in oligomers of various lengths. T4 DNA Ligase is dependent of ATP and works at room temperature while heating to 65°C can be used for deactivation.

3.2.7 Generation of plasmid constructs to obtain intein fusion proteins

Several CsgA-Intein fusion-constructs as well as HET-s-Intein fusion protein constructs were cloned and analyzed. The initial plasmids were: CsgA (21-152)-His expression plasmid based on pET11d (A.Ebert, HZI); HET-s(218-289)-His expression plasmid based on pET21 (C.Ritter, HZI); bacteriophage λ head protein D fused with *Npu* DnaE Intein domain N (gdD-IntN) plasmid based on pBAD and *Npu* DnaE Intein domain C fused with Thioredoxin (IntC-Trx) plasmid based on pRSF Duet (Vivien Schütz, Henning Mootz, Uni Münster). Cloning Vectors pET11a, pET28c, pET44a, pGEX6 and pETM20 were available from the HZI research group for recombinant protein expression (Joop van den Heuvel).

3.2.7.1 CsgA – intein fusion proteins with N-terminal His-Tag

The CsgA sequence coding for residues 21-69/89/130 were amplified by PCR from a CsgA (21-152) plasmid using the primers *Nde-6H-CsgA-for* and *CsgA-T70-rev* / *T90-rev* / *S131-rev,* the sequence for the Hexahistidine Tag was directly introduced in the primer. The sequence for the intein Domain N was amplified from the plasmid pAU08 using the primers *IntN-for* and *IntN-BamHI-rev*. The CsgA and IntN fragments were blunt end ligated and then reamplified using the primers *Nde-6H-CsgA-for* and *IntN-BamHI-rev.* The obtained fragment as well as a pET11a expression vector were digested with NdeI and BamHI and afterwards ligated using T4 DNA Ligase.

The corresponding sequences coding for CsgA residues 71/91/132-152 were amplified using the primers *CsgA-71/91/192-for* and *CsgA-BamHI-rev.* The sequence for the intein Domain C was amplified from the plasmid pVS01 using the primers *Nde-6H-IntC-for* and *IntC-rev.* This His Tag was directly introduced by the primer sequence. The IntC and CsgA fragments were blunt end ligated and then reamplified using the primers *Nde-6H-IntC-for* and *CsgA-BamHI-rev.* The obtained fragment as well as a pET11a expression vector were digested with NdeI and BamHI and afterwards ligated using T4 DNA Ligase.

3.2.7.2 CsgA-IntN constructs with C-terminal His Tag

The fusion constructs previously cloned into pET11a vectors were amplified by PCR using the primers *Nco-CsgA-for* and *IntN-Xho-rev.* The obtained fragment and a pET28c expression vector were digested with NcoI and XhoI and subsequently ligated. In this case the His-Tag is part of the pET28c vector and two additional residues arise through cloning with XhoI.

3.2.7.3 Solubility tagged IntCR5 constructs

Towards a fusion protein with NusA the sequence coding for IntCR5 was amplified by PCR from a pET11a plasmid using the primers *Bam-tev-IntCR5-for* and *IntCR5-rev.* The obtained fragment and a pET44a expression vector were digested using BamHI and HindIII and ligated using T4 DNA Ligase. Towards a fusion protein with Glutathion-S-tranferase (GST) the sequence coding for IntCR5 was amplified by PCR from a pET11a plasmid using the primers *Bam-tev-IntCR5-for* and *IntCR5-Xho-rev.* The obtained fragment and a pGEX6c expression vector were digested using BamHI and XhoI and ligated using T4 DNA Ligase. Towards a fusion protein with Thioredoxin (Trx) the sequence coding for IntCR5 was amplified by PCR from a pET11a plasmid using the primers *Nco-IntCR5-for* and *IntCR5-Xho-rev.* The obtained fragment and a pETM20 expression vector were digested using NcoI and XhoI and ligated using T4 DNA Ligase.

3.2.7.4 HET-s – intein Domain Fusion constructs

The HET-s fragment coding for residues 218-257 was amplified by PCR from a pET21 HET-s expression vector using the primers *Nde-hets-for* and *Hets-nterm-rev.* The IntN sequence has been amplified from the plasmid pAU08 using the primers *IntN-for* and *IntN-Xho-rev.* The PCR fragments were blunt end ligated and reamplified using the primers *Nde-hets-for* and *IntN-Xho-rev*. The obtained fragment as well as a pET21 expression vector were digested with NdeI and XhoI and ligated.
The IntC sequence has been amplified from the plasmid pVS01 using the primers *Nde-6H-IntC-for* and *IntC-rev*. The HET-s(258-289) sequence has been amplified from a pET21 Het-s(218-289) expression vector by using the primers *D258-cterm-for* and *cterm-rev*. Both fragments have been blunt end ligated and cloned into a pET11a expression vector using the NdeI and BamHI restriction sites

3.3 Biochemical Methods

3.3.1 Protein Expression and solubility tests

Small scale test expressions were performed to establish the recombinant expression conditions of novel gene constructs. Thus, 50-100 ml LB medium (+ Antibiotic) were inoculated with a single colony in Erlenmeyer flasks and cultures were grown to an OD600 of 0.6 – 0.8 at 37 °C. Optimal IPTG concentration, expression time and expression temperature were verified by varying these parameters. At several time points after induction samples were taken by V (µL) = 300/OD600 for SDS-PAGE analysis. The bacterial cultures were harvested by centrifugations and the cells resuspended in TRIS/NaCl buffer pH=8 and subsequently disrupted using a high pressure homogenizer. Insoluble material has been separated by centrifugation (16000 xg, 30 min) and both the supernatant and pellet were checked for protein overexpression by SDS-PAGE.

3.3.2 Recombinant Protein Expression

The exact conditions used for the recombinant expression of diverse constructs are listed in table 14.

Unlabeled
E.coli BL21 precultures were inoculated from an agar plate and incubated at 37°C over night. The main culture, usually 1L LB or TB medium in 3L Fernbach flasks, was inoculated with an appropriate volume of a preculture to reach a start OD600 of < 0.1. This main culture was grown to exponential phase (OD600 0.6 – 0.8) at 37 °C and shifted to optimal expression temperature. Gene expression was induced by adding IPTG to a final concentration of 0.1 – 1 mM. Cells

were harvested after 2 – 20 h induction by centrifugation (6000 g, 10 min, 4 °C) and further processed or stored at -20 °C.

Labeled

E.coli BL21 precultures were inoculated from an agar plate and incubated at 37°C over night. The first main culture, usually 1L LB medium in 3L Fernbach flasks, was inoculated with an appropriate volume of a preculture to reach a start OD600 of < 0.1. This main culture was grown to exponential phase (OD600 0.6 – 0.8) at 37 °C and the harvested by centrifugation. The cells were then resuspended in 1L CN040 Minimal medium containing 1g $^{15}NH_4Cl$ for nitrogen labeling and additionally 4 g ^{13}C-glucose for uniformly labeling. The bacterial cultures were then adjusted to the optimal expression temperature and induced with IPTG. Cells were harvested after 2 – 20 h induction by centrifugation (6000 g, 10 min, 4 °C) and further processed or stored at -20 °C

Table 14: Bacterial culture conditions for recombinant protein expression.

Gene Product	Antibiotic	IPTG Conc.	Induction Temperature (°C)	Induction Time (h)
CsgA	Ampicillin	1 mM	37	2
His-CsgA*-IntN	Ampicillin	0.2 mM	20	18
His-IntC-CsgA*	Ampicillin	0.5 mM	30	5
CsgA-IntN-His	Kanamycin	0.2 mM	20	18
HETs-IntN-His	Ampicillin	0.3 mM	20	18
His-IntC-HETs	Ampicillin	0.3 mM	20	18
Trx-IntC-R5	Ampicillin	0.5 mM	37	3

*) corresponding parts of CsgA fused with intein Domain were all expressed the same way

3.3.3 Cell lysis and purification of wt-CsgA under denaturing conditions

The complete cell pellets were therefore resuspended in denaturing buffer A (8 M GuHCl, 50 mM KPho, 150 mM NaCl, pH 7.2) and stirred 48 hours. Sonication (30min, 1s on / 4s off, Amplitude 70%) was applied to the cell suspension to affirm the complete dissociation and disrupt the DNA. The cell debris was removed by centrifugation (30 min.36000 xg) and the supernatant incubated (≥ 30min.) with equilibrated Ni^{2+}-Sepharose (1ml/l culture). Subsequently, the solid phase was packed into a column and immobilized proteins were washed threefold with 25 column volumes of buffer A containing 5, 10 and 20 mM imidazole. Finally proteins were eluted with buffer A containing 300 mM (3 x 5 CV) and 500 mM (5 CV) imidazole. Expression yields and purity of protein solutions were determined by SDS-PAGE. Elution fractions of high yield and purity were combined.

3.3.4 Cell lysis and Protein purification of intein fusion proteins and HET-s (218-289) under denaturing conditions

Every construct was either partially or completely expressed into inclusion bodies. The complete cell pellets were therefore resuspended in denaturing buffer B (6M GuHCl, 50 mM TRIS, 150 mM NaCl, pH 8) and stirred over night. Sonication (30min, 1s on / 4s off, Amplitude 70%) was applied to the cell suspension to affirm the complete dissociation and disrupt the DNA. The cell debris was removed by centrifugation (30 min.36000 xg) and the supernatant incubated with equilibrated Ni^{2+}-Sepharose (≥ 30min.). Afterwards the sepharose beads with bound target protein were collected by gentle centrifugation (2 min. 1000 xg) and transferred to a gravity flow column. The beads were washed with 10 column volumes buffer B containing 10mM imidazole and the target protein eluted with 20 ml buffer B containing 300 mM imidazole. Expression yield and purity of protein solutions were determined by SDS-PAGE and UV absorption measurements.

3.3.5 Refolding and protein *trans*-splicing of intein fusion proteins

Refolding of denatured proteins was initiated by dialyzing against a >100x fold access of native Buffer C (50 mM TRIS, 150 mM NaCl, 1mM EDTA pH 7.8). Minor precipitations were removed by centrifugation (15 min. 5000 xg). In case of HET-s intein fusions either dialysis against buffer C containing 1M GuHCl or buffer exchange over a column was applied.

Refolded splicing partners (e.g. R1IntN + IntCR25) were mixed equimolarly and diluted with dialysis buffer to keep the protein concentrations during the splicing reaction under 20 µM. The splicing reaction was initiated by the addition of 0,5-1 mM TCEP (*tris*-(2-carboxyethyl)-phosphine). Large scale experiments were conducted in Falcon tubes at room temperature and gentle agitation for 48 hours. Samples for SDS-PAGE were taken at several time points to follow the splicing efficiency. The precipitated proteins were collected by centrifugation (15 min. 6000 xg). In the initial sample preparation strategy of segmentally labeled CsgA the resulting pellet was washed several times with 5-6 M GuHCl buffers and loaded to a solid state NMR rotor. The optimized sample preparation protocol that has been applied to CsgA and HET-s requires the complete dissolution of the protein pellet. Aggregates of HET-s(218-289) and the corresponding intein fusion proteins are soluble in 6M GuHCl. Protein pellets that arose from splicing towards segmentally labeled CsgA could be resolved in 5M guanidine thiocyanate Buffer (5 M GuTC, 50 mM TRIS, 150 mM NaCl, pH 8).

3.3.6 Reversed Affinity Chromatography

The term “reversed” in this case indicates that the protein of interest should not bind to the column and gets purified from affinity tagged proteins. This technique was exploited to purify tag-free splicing product, CsgA or HET-s, from precipitated educts or side products. Around 5 ml of Ni^{2+}-Sepharose were equilibrated with 6 M GuHCl, 50 mM TRIS, 150 mM NaCl, 10 mM β-mercaptoethanol, pH 8 and incubated in Falcon tubes with dissolved HET-s splicing products. CsgA splicing products, dissolved in 5M GuTC were diluted 1:2 with 6M GuHCl, adjusted to 10 mM β-mercaptoethanol and incubated with equilibrated Ni^{2+}-Sepharose. The solution was passed through a gravity flow column and the flow through collected; one column volume of GuHCl buffer was added to minimize the target protein loss.

3.3.7 Size Exclusion Chromatography

Dissolved molecules are separated according to their size, based on the different permeation in a carrier material with a controlled pore size. The hydrodynamic radius is the critical parameter in the separation process. Molecules from a certain size cannot enter the pores and elute in the void volume. Smaller molecules penetrate the pores and thus experience a delay in the elution volume. The mobile phase serves only as a solvent and has no direct influence on the separation. After calibration with different sized substances the molecular weight of the analyzed protein can be estimated. Here the column was operated by an automated chromatography device, Äkta Explorer (Ge Healthcare) under denaturing conditions using 6 M GuHCl, 50 mM TRIS, 150 mM NaCl, 1 mM DTT pH 8 as mobile phase. A Superdex 200 16/60 column was used to purify monomeric CsgA.

3.3.8 Amyloid formation – *in vitro* fibrillization

Fibrillization of CsgA or HET-s monomers was achieved by fast buffer exchange to native buffer lacking any denaturing agent. Desalting columns are based on a Sephadex G-25 stationary phase and commercially available for sample volumes from 0.5 – 15 ml. Recombinantly expressed CsgA (21-152) was desalted to a Phosphate Buffer pH 7.2 using a self-packed column and incubated under gentle agitation at room temperature. Segmentally labeled CsgA samples were desalted to 50 mM KPho, 1 mM DTT, 0.05 % NaN_3 over a Hiprep Desalting 26/10 column (Ge Healthcare). HET-s samples were concentrated to final sample volume of 2.5 ml using a Amicon centrifugal filter (Millipore) and desalted to 50 mM TRIS, 150 mM NaCl pH 7.2 using a PD-10 gravity flow desalting column (Ge Healthcare). Fibrilization occurred automatically by incubating over night.

3.4 Protein analytical Methods

3.4.1 Physico-chemical parameters of the studied proteins

The molecular weight (MW), the isoelectric point (pI) and the molecular absorption coefficient (ε) of every studied protein was calculated based on the primary sequence using the ProtParam tool (http://web.expasy.org/protparam/). The parameters are listed in table 15.

Table 15: Biophysical parameters of proteins used in this study.

Protein	Number of residues	MW (Da)	pI	ε_{280} (M^{-1} cm^{-1})
CsgA (21-152)	138	14050	5.73	11460
R1IntN	159	17740	4,83	17545
R12IntN	179	19595	4,7	17545
R14IntN	221	23950	4,69	23045
IntCR25	123	13103	6,81	9970
IntCR35	105	11496	8,86	9970
IntCR5	63	7141	9,36	4470
Trx-IntCR5	190	20698	6,34	20065
GST-IntCR5	294	33863	6,26	49070
Nus-IntCR5	628	69157	4,91	37150
HETs218-257-IntN	150	17204	5,41	13075
IntC-HETs258-289	75	8491	7,97	9970

3.4.2 Protein concentration determination by UV/VIS-spectroscopy

UV / VIS spectroscopy is an electron spectroscopy, meaning that electrons- more precisely, valence electrons – are excited with electromagnetic waves of visible and ultraviolet light. According to the molecule orbital theory, overlapping orbitals are divided into bonding and anti-bonding orbitals. In the ground state all electrons are in the bonding orbital and the anti-bonding orbital remains empty. Upon excitation, electrons change to the anti-bonding orbital. The return to the ground state occurs without radiation. The absorption of electromagnetic radiation with a predetermined energy, or wavelength, is described by the Lambert- Beer law, assuming a homogeneous distribution of the absorbing substance.

$$A = \log\left(\frac{I_0}{I}\right) = \epsilon * c * d$$

Where I_0 and I represent the light intensity before and after permeating the sample, c is the concentration of the absorbing material, and d is the diameter of the solution or the cuvette. The constant ε is called the molar absorption coefficient which is protein dependent and can be calculated based on the amino acid sequence. For the measurement of protein concentrations a wavelength of 280 nm is used which corresponds to the absorption maximum aromatic amino acids, especially Tryptophan.

3.4.3 NaDOC/TCA Precipitation

In preparation for SDS-polyacrylamide gel electrophoresis proteins from guanidine hydrochloride containing solutions had to be precipitated by sodium deoxycholate and trichloracetic acid (Bensadoun and Weinstein 1976). Corresponding samples were diluted 1:10 in 0.1 % (w/v) NaDOC and mixed with TCA to a final concentration of 10 % (w/v). After incubation for 30 min at 4 °C the samples were spun down (20800 g, 30 min, 4 °C). The obtained pellets were washed twice by adding 500 µL chilled acetone, which was removed after centrifugation (20800 g, 10 min) and the pellets were air-dried to remove retained acetone. Dried pellets were dissolved in 20 µL SDS sample buffer and applied on a SDS gel.

3.4.4 Polyacrylamide Gel Electrophoresis

SDS - polyacrylamide gel electrophoresis is a method to separate proteins according to their molecular weight which is used to determine the amount and purity of proteins. SDS (sodium dodecyl sulfate) is an anionic detergent, which covers the inherent charge of the proteins so that the separation in the gel is only dependent of the molecular weight. Proteinaceous samples are prepared by adding a sample buffer and incubating at 95°C whereas all proteins are denatured and reduced and can readily loaded on a gel.

Polyacrylamid gels are chemically inert and highly stable. They are produced by copolymerization of acrylamid monomers, ammonium persulfate (APS) and Tetramethyl-ethylenediamine (TEMED). Commercial NuPAGE Novex 4-12% Bis-Tris Midi Gels (Invitrogen, 1.0 mm x 26 well) were used at 200V for 45 min. If not stated otherwise, 10 µL samples were loaded per lane. Finally, the gels were stained with Coomassie blue containing solution (Instant Blue, Gentaur, Belgium).

3.4.5 Western Blot

Western blot is a technique to detect proteins, previously separated by SDS-PAGE, with a specific antibody. The protein samples from SDS-polyacrylamid gels were immobilized on polyvinylidene difluoride (PVDF) membranes (Immobilon P, Millipore). First, the membrane was activated in methanol (~ 10 s) and washed with transfer buffer. The gels were equilibrated in transfer buffer as well (15 min). After that, the gel was placed onto the membrane, both placed between two layers of soaked Whatman paper onto the anode of a semi-dry blot apparatus. The transfer was carried out for 30 min at 12 V.

After western blotting unspecific binding sites on the PVDF membrane were saturated for 1 h in blocking solution and then washed in TBST (5 min). The antibodies were diluted in blocking solution. The membrane was incubated with Anti-His antibody for 4 hours and after washing with TBST (3 x 5 min) additionally incubated for 2 h with an AP conjugated secondary antibody. Finally, the membrane was washed (TBST, 3 x 5 min, AP buffer 1x 5min) and immobilised proteins were detected on the membrane by the AP catalyzed reaction of BCIP/NBT (Promega).

3.4.6 N-Terminal Sequencing

Proteins, which were transferred to a membrane as described before, were excised after visualizing with Ponceau S staining solution and sequenced by automated Edman degradation (Edman and Begg 1967). N-terminal sequencing was carried out by Rita Getzlaff or Beate Jaschok-Kentner (HZI, Braunschweig) using a 494A HT Protein Sequencer (Applied Biosystems).

3.4.7 Mass Spectrometry

Matrix assisted laser desorption ionization (MALDI) mass spectrometry was used to verify the mass and intactness of samples produced by protein trans-splicing. Aggregated samples (approximately 25 - 50 µg) were dissolved in 98 % (v/v) FA, loaded onto a prespotted anchor chip (PAC) target with an α-cyano-4-hydroxycinnamic acid matrix and dried at RT. To avoid N-formylation the FA containing solutions were processed as fast as possible. The molecular masses were determined in the positive-ion mode on a Bruker Ultraflex time-of-flight mass spectrometer (Bruker Daltonics GmbH). All experiments were performed by Anja Meier and Undine Felgenträger (HZI, Braunschweig).

3.5 Protein Structure Analysis

Secondary Structure Analysis

CD experiments and FT-IR spectroscopy experiments were performed to probe the secondary structure content of a protein. Both techniques can be applied to samples in solution as well as to protein aggregates. They are fast, quantitative and non-destructive photometric techniques providing detailed information without time intense data interpretation .

3.5.1 Circular Dichroism

CD spectroscopy is a method to study the optical activity of asymmetric molecules in solution, more precisely to analyze is the absorption difference of circularly polarized. Optical activity occurs for instance by the introduction of a chiral center in a molecule. The optically active substance has different refractive indices for left-and right-polarized light and therefore different absorption values. For all chiral molecules, the difference A_{Left} - A_{Right} is unequal to zero, this behavior is called circular dichroism. In CD experiments the differential absorption of left and right circular polarised light by chiral molecules is recorded. Secondary structure can be determined by CD spectroscopy in the "far-UV" spectral region (190-250 nm). At these wavelengths the chromophore is the peptide bond, and the signal changes when it is located in

a regular, folded environment. α-helix, β-sheet and random coil structures each give rise to a characteristic CD spectrum like illustrated in figure 17.

In preparation of CD experiments, protein samples were first transferred to a CD suitable buffer (50 mM KPi, 150 mM NaF, pH 7.2) by desalting of soluble proteins or washing of specific aggregates. If not stated otherwise, the protein concentration was adjusted to 1 mM peptide bonds and samples were applied to cuvettes with a path length of 1 mm (Hellma). Spectra were measured using a JASCO J-810 CD-spectrometer. Measurements were performed in far UV (190 – 260 nm) range in continuous scanning mode at 20°C.

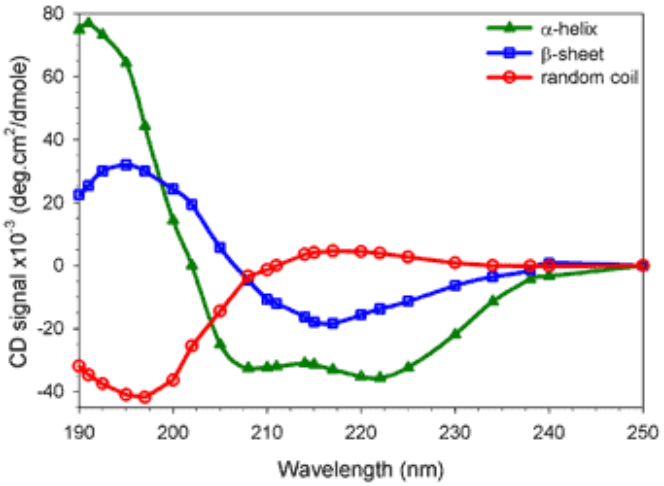


Figure 17: Reference CD spectra of α-helical, β-sheet or random coiled proteins (reference: http://www.fbs.leeds.ac.uk/facilities/cd/).

3.5.2 Fourier Transform Infrared Spectroscopy

FTIR spectroscopy is a measurement of wavelength and intensity of the absorption of IR radiation by a sample. It is based on measuring molecular bond vibrational frequencies. The polypeptide and protein repeat units give rise to nine characteristic IR absorption bands, of which the amide I and II bands are the two most prominent vibrational bands of the protein backbone. An amide I absorption maximum at a wavenumber of ~1650 cm^{-1} indicated α-helical content while a maximum of ~1630 cm^{-1} is characteristic for β-sheet. In case of an amyloid characteristic cross-β structure, the maximum is shifted to ~1620 cm^{-1}.

FTIR spectroscopy was carried out using a Bruker Tensor 27 system (Bruker Optics) equipped with a BioATR cell (ZnSe single crystal), a mid-IR source (4000 – 400 cm-1) and a KBr beamsplitter. The protein samples were applied as fine suspension with a protein concentration of 1 – 2 mg/mL. In preparation of the FTIR experiments, fibrils were washed three times with water by repeating centrifugation (20800 g, 5 min, RT) and re-suspension steps. Finally fibrils were re-suspended and sonicated for 10 min in an ultrasonic bath. Measurements were carried out in absorbance mode using following conditions: sample and background scan time was 200 scans, optical filter was open and aperture was set to 6 mm, detector setting was given by LN-MCT Photovoltaic and preamp gain A.

3.5.3 Solution NMR Spectroscopy

Solution NMR was used to confirm the labeling pattern of ^{15}N segmentally labeled CsgA samples. Experiments were carried out at 25 °C using a Bruker Avance III 600 spectrometer equipped with four radio-frequency channels and a 5 mm Z-axis gradient triple-resonance cryo-probehead. The offset that represses the water signal and the pulse length were optimized for every individual dataset. Amyloid fibrils were collected by centrifugation (20800 g, 4 min, RT) and solubilised in ~ 500 µL aprotic DMSO (perdeuterated, d6-DMSO) containing up to 0.1 % (v/v) TFA (deuterated, d1-TFA), and placed in a 5 mm NMR tube (Norrel). For 1D- and 2D-experiments ~ 20 mg wet pellet of fibrils was processed. Heteronuclear single quantum coherence (HSQC) spectra were acquired and processed using Topspin 2.1 (Bruker Biospin) and analyzed by CCPN 2.1 (Vranken, Boucher et al. 2005). In HSQC spectra every amide proton is correlated with the bound amide nitrogen (Ernst, Bodenhausen et al. 1987, Cavanagh 2007). The evaluation of the labeling pattern was based on a backbone assignment of CsgA (21-152) that has been performed in our group during an earlier study.

3.5.4 Solid state NMR Spectroscopy

For solid state NMR all CsgA and HET-s samples were centrifuged at 100000 x g for 30 minutes and transferred to ZrO rotors at 21000 x g using a table to centrifuge. All solid state NMR spectra were recorded using 3.2 mm rotors and the temperature was kept constantly at 278K. The experiments were acquired on a 14.1 T Bruker Advance III standard bore spectrometer (^1H Larmor frequency of 600 MHz) equipped with a 3.2 mm TriGamma CP/MAS probe, located at the HZI Braunschweig; a 14.1 T Bruker Advance wide bore spectrometer equipped with a 3.2 mm standard CP/MAS probe or on a 21 T Bruker Advance standard bore spectrometer (^1H Larmor frequency of 900 MHz) equipped with a 3.2 mm standard CP/MAS probe, both located at the FMP Berlin. Heteronuclear decoupling was applied using SPINAL64 on protons with powers of 90-100 kHz. Cross polarization with contact times of 1400-1800 ms were used to transfer magnetization between nuclei. All important parameters regarding the data acquisition of the different samples are summarized in table 16.

Table 16: Solid State NMR Data recorded during this work.

Sample	Experiment	Magnetic field strenght	Spinning frequency	Acquisition time (aq1/aq2/aq3)	DARR Mixing time
CsgA (21-251)	2D DARR	21 T	13 kHz	15 ms 4 ms	50 ms
R1 CsgA	2D DARR	14.1 T	10 kHz	14 ms 5 ms	25 ms
	2D DARR	21 T	13 kHz	15 ms 3.5 ms	50 ms
	2D DARR	21 T	13 kHz	15 ms 3.2 ms	300 ms
	2D NCA	21 T	13 kHz	15 ms 9.6 ms	-
	2D NCACX	21 T	13 kHz	15 ms 9.6 ms	25 ms
	3D NCACX	21 T	13 kHz	12 ms 5 ms 6 ms	25 ms
	2D NCOCX	21 T	13 kHz	15 ms 9.6 ms	25 ms
R 1+2 CsgA	2D DARR	14.1 T	10 kHz	11.9 ms 5.3 ms	50 ms
HET-s(218-289)	2D DARR	14.1 T	10 kHz	11.9 ms 5.6 ms	50 ms
	2D DARR	14.1 T	10 kHz	11.9 ms 5.6 ms	250 ms
HET-s(^{13}C-218-256)	2D DARR	14.1 T	10 kHz	11.9 ms 6.2 ms	50 ms
	2D DARR	14.1 T	10 kHz	11.9 ms 3.9 ms	200 ms
HET-s(^{13}C-257-289)	2D DARR	21 T	13 kHz	20 ms 3.5 ms	50 ms

All spectra were measured and processed using Topspin 2.1 (Bruker Biospin). Data analysis, assignments and determination of secondary structure was achieved using CCPN 2.1 (Vranken, Boucher et al. 2005).

4 RESULTS

4.1 STRUCTURAL INVESTIGATION OF CSGA FIBRILS

CsgA is the major subunit of a bacterial functional amyloid called curli, expressed by enterobacteriaceae like *Escherichia Coli*. Curli are involved in formation of biofilms and host cell attachment. CsgA consists of 152 residues starting with a Sec export signal which gets cleaved upon translocation to the periplasm, followed by 22 flexible amino acids and 5 imperfect repeats, R1-R5, with an average length of 20 residues. Amyloids of CsgA have been extensively studied *in vivo* as well as *in vitro* and provide a well suited system for understanding of amyloid formation, propagation and control mechanism. Highly resolved structures of amyloid fibrils are required to get insights in functions and mechanisms as well as differences between different types of amyloids, such as disease related and functional ones. Solid state NMR has evolved to be a powerful tool for the elucidation of insoluble protein structures. Here, *in vitro* fibrillized CsgA was studied with solid state NMR spectroscopy. For this, protein trans-splicing was adjusted for the use with insoluble samples in order to label just defined regions of CsgA.
A stable system for recombinant expression and purification of CsgA 21-152 was established in a previous study. The construct lacks the Sec export-signal sequence; therefore, CsgA was expressed in inclusion bodies and purified under denaturing conditions. The yields from minimal medium expression cultures (~10mg/L culture) were high enough for structural studies and, after the optimization of fibrillization conditions, homogeneous samples could be obtained. The exact location of β-strands was identified by an extensive H/D exchange solution NMR analysis.

In the present study the low resolution data should be complemented with high resolution solid state NMR investigations. Only rigid domains of proteins give rise to ssNMR cross peaks. Based on H/D exchange Data, at least 90 spin systems were expected to be detectable by solid state NMR. Therefore, uniformly labeled CsgA was produced following the established protocol (see chapter 3.3). Initial ssNMR spectra comprised a decent signal to noise ratio, but 2D ^{13}C-^{13}C correlation spectra appeared to be relatively empty regarding the number of expected spin systems and compared to spectra of uniformly labeled HET-s(218-289) (see figure 11). Many efforts were carried out to improve the sample and the spectral quality such as optimizing the *in vitro* fibrillization conditions. However, after careful reproduction, the solid state NMR data was proven to be reliable and the contribution of all expected amino acids could be verified at several sections of the spectrum after thorough examination. The aliphatic region of a representative 2D ^{13}C-^{13}C spectrum is illustrated in figure 18 a. The Cγ1-Cδ1 crosspeaks of the three expected isoleucine residues, highlighted in figure 18 b, showed a narrow linewidth of less than 1 ppm which is comparable to published data of amyloid fibrils. No indications about the presence of more than a single fibrillar isoform such as peak broadening or the appearance of more than one set of NMR resonances were detected. The evaluation of the well separated Cα-Cβ cross peaks of Thr, Ala, Ile and Val confirmed the overall β-sheet structure of CsgA.

The severe shortcoming of the spectra was the strong spectral overlap in almost every region, highlighted in figure 18 c, where all 8 threonine Cβ-Cα correlations appeared in a single, large peak. This property hampered the sequential NMR assignment of uniformly labeled CsgA fibrils which is a critical step towards structure investigations. Additionally, even if a backbone assignment would be successfully conducted, the assignment of unambiguous distance restraints would be highly challenging. Therefore an isotope labeling strategy, so far just described for soluble proteins, was exploited to get unambiguous assignments of CsgA.

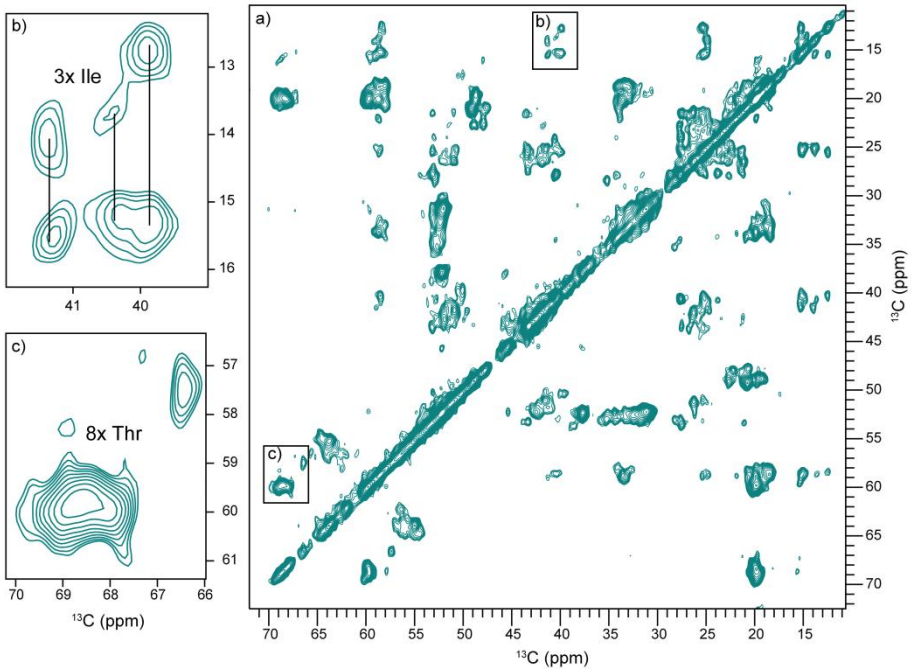


Figure 18: 2D DARR spectrum of CsgA (21-152) amyloid fibrils measured at 21T, ω_r of 13,5 kHz and with a mixing time of 50 ms. The complete aliphatic region is shown in a), threonine and isoleucine cross peaks are highlighted in b) and c)

4.1.1 Segmental Isotope Labeling of CsgA Amyloid Fibrils

Segmental isotope labeling by protein trans-splicing (PTS) was used to overcome this challenge. As explained in the introduction, PTS involves the use of a split intein which is able to reassemble to a functional state and ligate the connected exteins by a peptide bond. By this, just a domain of a protein can be labeled with isotopes while the rest of the protein remains unlabeled. The DnaE intein from *Nostoc punctiforme* was chosen due to its high splicing activity even under harsh conditions and its high acceptance of flanking amino acids at the splicing junction. Therefore, several complementary pairs of fusion proteins containing parts of CsgA linked to an intein Domain, IntN and IntC, were generated. Threonine or serine residues were chosen to be replaced with the catalytic cysteine of the intein and no further linker residues were introduced at the splicing junction in order to produce a protein biochemically and biophysically close to the wild type one. Blunt end ligation of PCR fragments was used to avoid additional residues from restriction side base pairs. As a result, the splicing product comprised only a cysteine point mutation. The biophysical similarity of cysteine mutants to wild type protein was already demonstrated for several examples, in an earlier study.

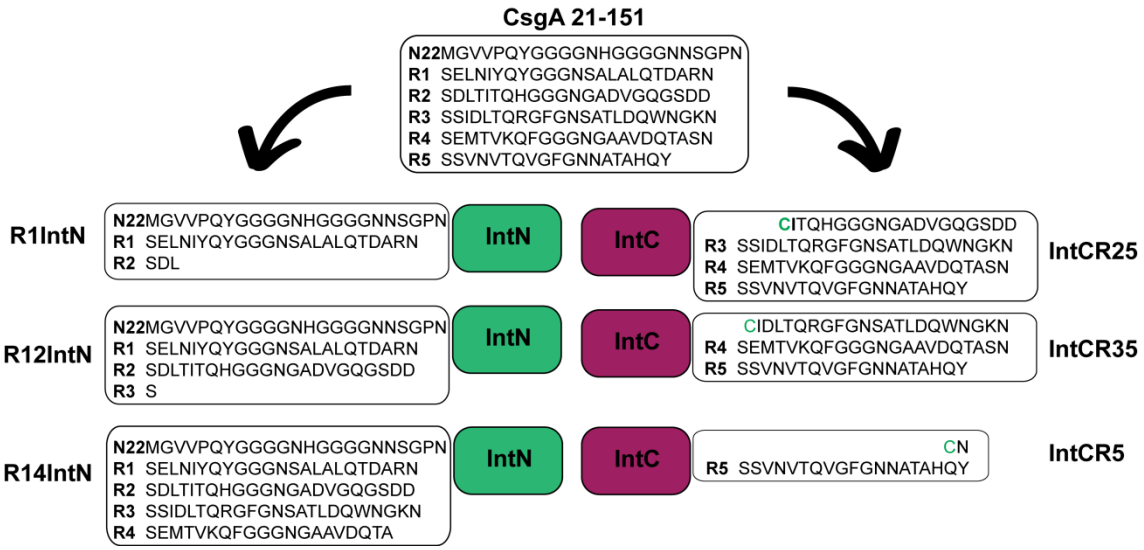


Figure 19: Design of split intein CsgA fusion proteins for segmental isotope labeling. Constructs containing the N-terminal intein domain were fused with CsgA residues 21-70 (R1IntN), 21-90 (R12IntN) or 21-131 (R14IntN). Corresponding residues for the production of full length CsgA (21-152) were fused with the C-terminal intein domain (IntCR25, IntCR35, IntCR5).

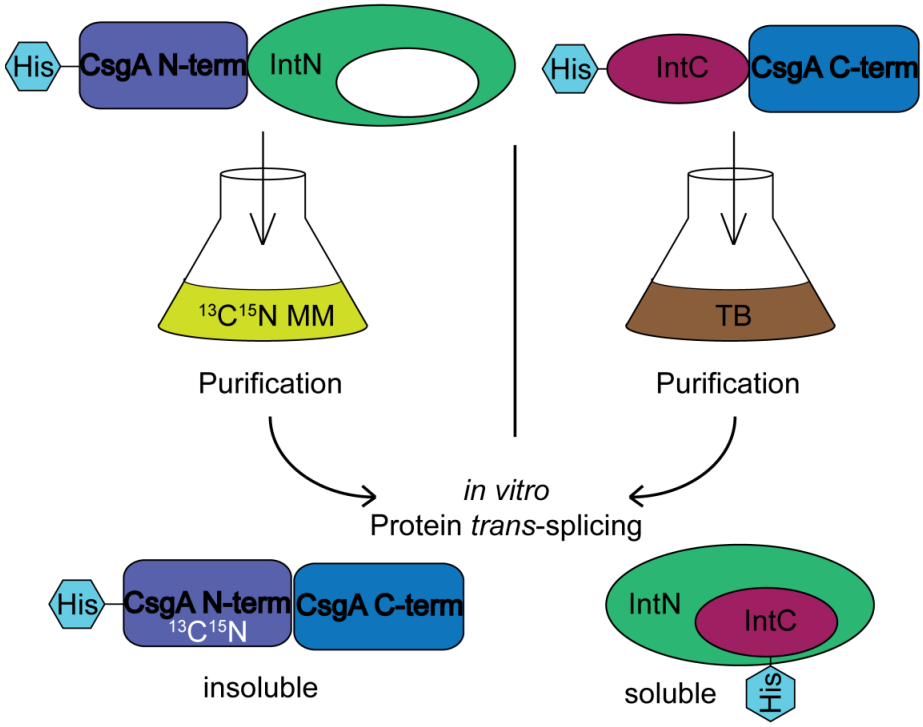


Figure 20: Initial strategy for segmental isotope labeling of CsgA amyloid firbils. Intein-CsgA fusion proteins were produced by recombinant expression in *E.coli*, one part in isotope labeled minimal media and the corresponding part unlabeled in TB media. After affinity purification over N-terminal His Tags, protein trans-splicing was initiated in vitro. The segmentally labeled CsgA precipitated during the reaction while the intein complex remained in solution and could be separated by centrifugation.

The selected positions to split the sequence were Thr70, Ser90 and Ser 132 which gave rise to label specific repeat units: either just R1, R1 and R2 (and/or the corresponding C-terminal segment) or R5 (and/or the corresponding N-terminal segment). Resulting IntN fusions contained an N-terminal segment of CsgA (21-69/89/131) while IntC fusions carried the corresponding C-terminal part to produce full length CsgA samples. The design of CsgA intein

fusions is illustrated in figure 19. Based on the containing CsgA repeats the fusion proteins were named e.g. R12IntN where 12 stood for CsgA repeats 1-2 and IntCR25 where 25 stood for repeats 2-5.

The segmental isotope labeling was achieved by recombinant expression of the fusion protein pairs, separated from each other; one part labeled with stable isotopes in minimal media and the corresponding part unlabeled in LB or TB medium. Suitable conditions for protein *trans*-splicing efficiency and fusion protein stability had to be adjusted. Several parameters were tested and optimized such as the protein concentration, the buffer conditions, the temperature and the type and concentration of the reductant.

Hexa-histidine (His) Tags were chosen for purification by affinity chromatography, due to their functionality under denaturing conditions. Since a single His-Tag connected to a split intein part was shown to be functional in a previously published study, N-terminal Tags were fused to CsgA N-term and IntC. The splicing product results in a single His-Tag labeled protein which matches the situation of the full length construct used to produce the uniformly labeled CsgA sample. The product purification strategy was based on the assumption that the splicing product (CsgA) precipitates while the educts and the by- products (intein complex) would remain soluble and could simply be removed by centrifugation. The resulting pellet would not need to be dissolved again and could directly be loaded to a solid state NMR rotor. The strategy is summarized in figure 20.

4.1.2 Expression and purification of CsgA - intein Fusions

The required amount of pure target protein for a solid state NMR sample is around 10 mg. Therefore, a high expression yield of natively folded precursor proteins from minimal media is mandatory, considering the high loss during protein splicing and product purification and the cost intensive ^{13}C labeling. It is a simple fact that amyloid fibrils formed by CsgA are extremely insoluble and it turned out that intein Fusion proteins were still poorly soluble. All CsgA-IntN fusion could be expressed partially soluble while IntC fusion appeared to express entirely in inclusion bodies. The small IntC domain (~30 residues) was not sufficient to keep the highly insoluble CsgA parts in solution while the larger IntN domain (~100 residues) contributed enough charged amino acids to increase the solubility. However, the purification was perfomed under denaturing conditions and the pure fusion proteins were refolded into a functional state by dialysis against a TRIS buffer without any denaturing agents. Just a minor protein fraction precipitated during refolding. The fusion proteins were stabilized by dimerization over a disulphide bridge (data not shown). As an example the purification of R1IntN and IntCR25 analyzed by SDS-PAGE is shown in figure 21 a: Unspecific binding was very limited under denaturing conditions and a purity of more than 95% could be estimated for both proteins. A weak additional band arose from a cleavage reaction due to the high amount of β-mercaptoethanol in the loading buffer.

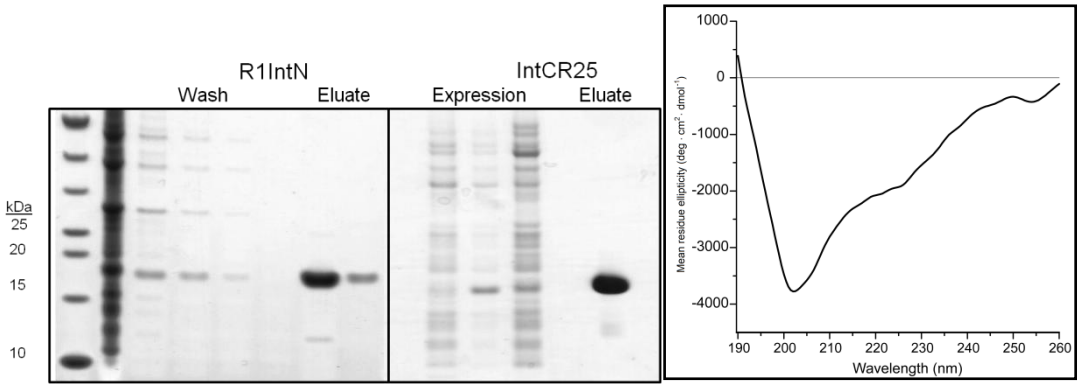


Figure 21: a) SDS-PAGE analysis of the recombinant expression and affinity purification of CsgA-intein fusion proteins. Pure protein can be eluted from the column, additional lower bands arise from a cleavage reaction caused from high concentrations of β-mercaptoethanol in the sample buffer b) CD-spectrum of 10 µM refolded IntCR25 showing mostly random coil character with a small percentage of β-sheet

The expression yields from one liter of minimal media after purification and refolding were: R1IntN 25 mg, R12IntN 30 mg, R14IntN 50 mg, IntCR25 10 mg, IntCR35 8 mg and IntCR5 4 mg. Generally, the IntN fusion protein expression yields were higher than IntC. Towards a NMR sample, R1IntN and R2IntN yielded sufficient amounts while the expression level of IntCR5 appeared to be very low. Additionally, the analysis of IntCR5 turned out to be challenging due to the small size and low extinction coefficient.

Following purification and dialysis, protein folding was determined by CD-spectroscopy, an example spectrum of IntCR25 is shown in figure 21 b. A correct fold of the intein domains is required for trans-splicing activity while folding of the CsgA part would most probably lead to aggregation and after splicing to an unpredictable starting condition of the fibrillization. For IntCR25 a β-sheet content of >80% was expected in case of refolding of both the IntC and CsgA domain and 18% if only the intein domain adopted a folded structure. The obtained spectrum showed a relatively high amount of random coil and a bit of β-sheet, indicating that a proper refolding of the inteinC fragment (34 residues, 21 β-sheet) occurred while the CsgA peptide (82 residues) remained unfolded. It could be concluded that the requirements for protein trans-splicing were fulfilled.

4.1.3 IN VITRO PROTEIN TRANS-SPLICING

For protein *trans*- splicing, the precursor proteins were mixed equimolarly and incubated several days. Low protein concentrations (<20µM), ambient temperature, TRIS buffers at pH 7.8, TCEP as reductant and a continuous slow agitation resulted in the highest splicing efficiency which could be estimated from band intensities of the intein fragments on SDS-PAGE gels and roughly from the amount of precipitated protein. The splicing yield was significantly decreased at high concentrations due to aggregation of the precursor proteins. The protein stability was increased at lower temperatures (4°C) or by additions of 2M Urea or 1M GuHCl but at the same time the splicing efficiency reduced and therefore the final yield was not improved. The use of DTT as a reductant resulted in high amount of intein cleavage caused by its nucleophilic character.

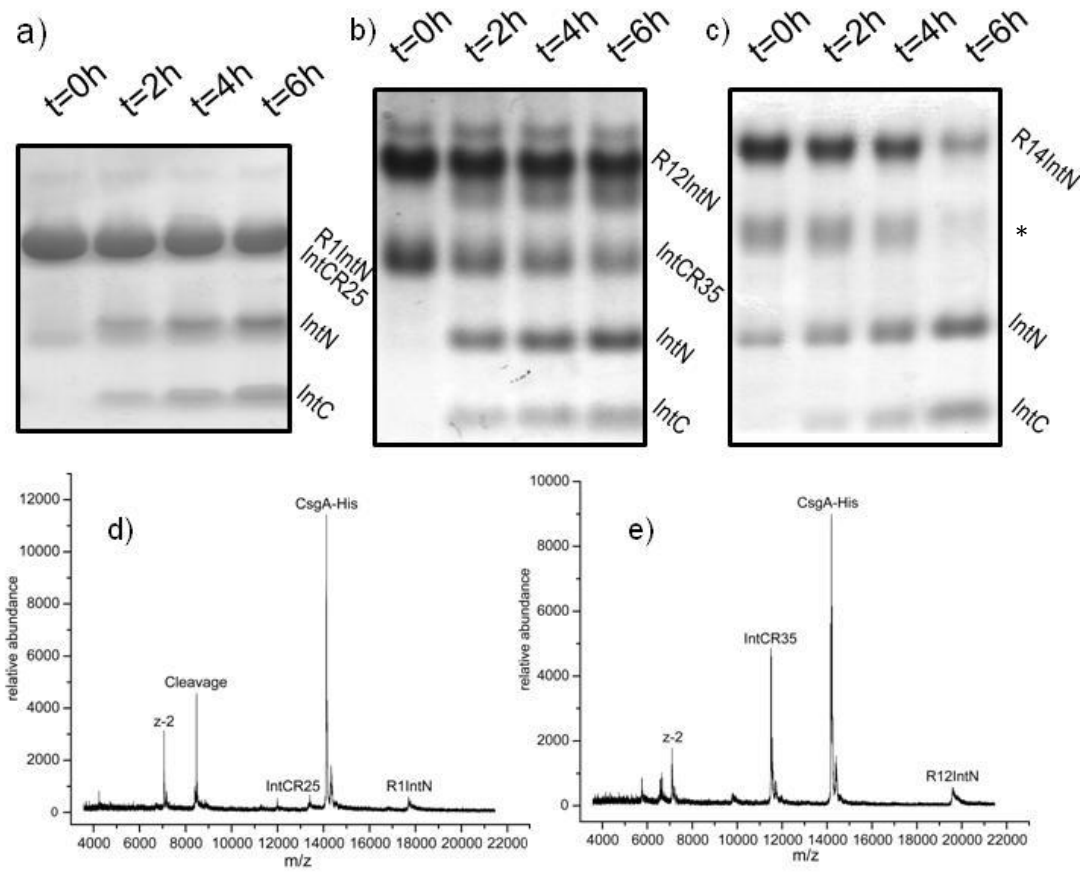


Figure 22: *In vitro* protein trans-splicing of CsgA-intein fusion proteins: a) R1IntN + IntCR25, b) R12IntN + IntCR35, c) R14IntN + IntCR5. Samples for SDS-PAGE were taken after 0, 2, 4 and 6 hours of incubation. The intein domains appear over time while the splicing product CsgA is not soluble in SDS-PAGE sample buffer. Contaminations are marked with an asterisk. d) MALDI-MS spectrum of the resulting protein pellet of a splicing reaction of R1IntN and IntCR25. CsgA-His showed the major peak, minor educt leftovers and a mass resulting from an intein cleavage reaction could be detected. e) MALDI-MS spectrum of the resulting protein pellet of a splicing reaction of R12IntN and IntCR35. CsgA-His showed the major peak, an intensive peak was detected for precipitated IntCR35 as well as a minor peak of R12IntN

All three pairs of refolded precursor proteins showed splicing activity but also precipitated during the splicing reaction. SDS-PAGE samples taken after various time-point of protein splicing of a) R1IntN + IntCR25, b) R12IntN + IntCR35 and c) R14IntN + IntCR5 are shown in figure 22. The fusion proteins R1IntN and IntCR25 permeated the gel at exactly the same position and IntCR5 was not detectable. The band intensity of intein domains increased over time indicating the ongoing reaction progress. The efficiency was estimated to be around 30-40%, based on SDS-PAGE analysis. The precipitate could be analyzed by electrophoresis and mass spectrometry after solubilization with formic acid. The expected trans-splicing product (CsgA-His) could be determined by MALDI-mass spectrometry for all three constructs. Figure 22 d) shows the MALDI-MS analysis of the precipitated proteins during a trans-splicing reaction of R1IntN and IntCR25. The splicing product, His-tagged CsgA, was detected as the major peak; furthermore the spectrum comprised minor peaks from both intein fusion precursor proteins and from the intein cleavage product R25. CsgA was also detected in the protein pellet that arose from a splicing reaction of R12IntN and IntCR35. Figure 22 e) shows that the peaks of precipitated educts were more intense in this case but less cleavage products were detected. The precursor

pair for the R1 sample showed the highest stability and the best splicing reproducibility and was therefore chosen for up scaled experiments.

4.1.4 Analysis of the splicing product

The resulting protein aggregates from a splicing reaction of 10 ml, 15 µM R1IntN and IntCR25 was collected by centrifugation, washed and subsequently analyzed in terms of purity and biophysical properties. The SDS-PAGE in figure 23 a) showed a strong band at 19 kDa, which corresponded to CsgA. The intein complex was not detected on the gel and therefore stayed in solution and was removed by centrifugation. However, not a clear single band but additional minor contaminations with a slightly lower molecular weight remained.

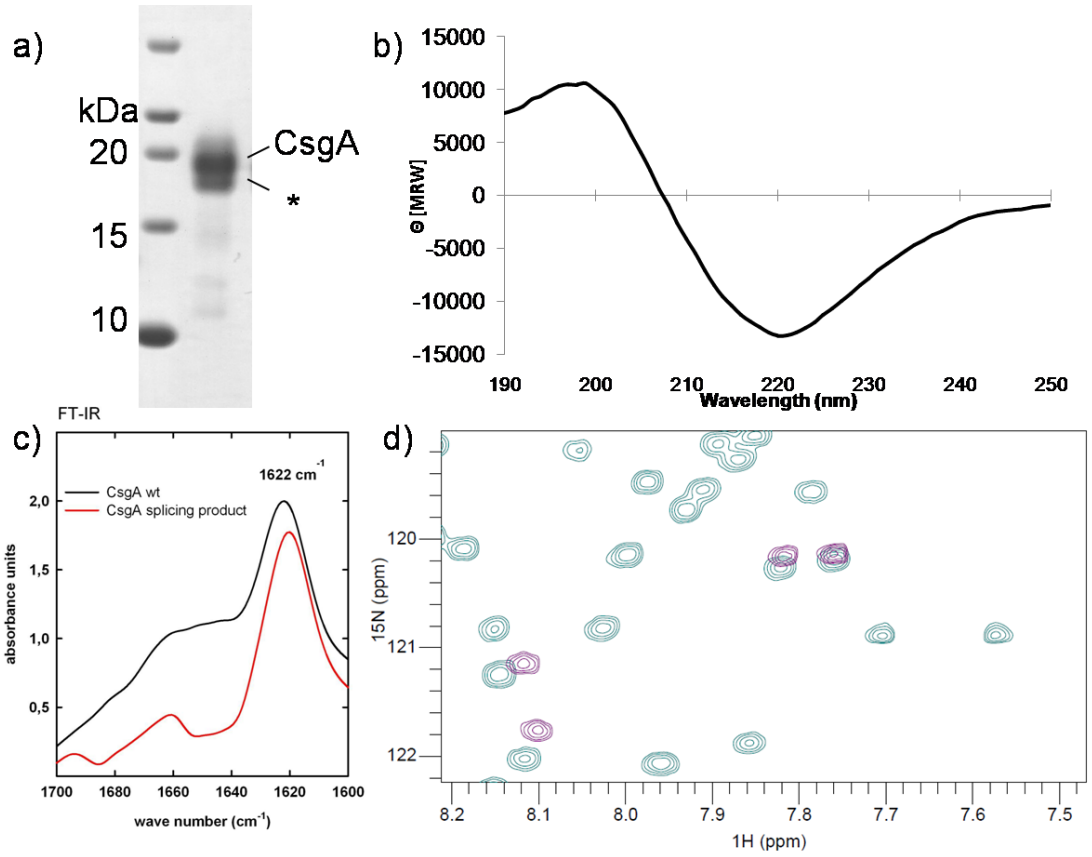


Figure 23: Analysis of the splicing product. a) the sample was dissolved in formic acid, dried and analyzed by SDS-PAGE. CsgA and some minor contaminations could be detected. b) CD-spectroscopy of the precipitated protein showed an all beta spectra, c) the amyloid fold was confirmed by FTIR spectroscopy showing the amyloid specific maximum comparable to wild type CsgA d) a ^{15}N labeled sample was dissolved in DMSO and the labeling pattern confirmed by solution NMR. The overlay of ^{1}H-^{15}N HSQC acquired on uniformly labeled CsgA and segmentally labeled CsgA shows a reduced number of peaks while maintaining the peak positions.

Furthermore, the sample was analyzed by CD- and FTIR- spectroscopy to get more information about the fold and amyloid characteristics of the sample. The CD-spectra showed a pure β-sheet structure, like expected for CsgA amyloid fibrils (Fig. 23 b). In comparison with wild type CsgA fibrils, the fibrils produced by protein splicing showed the same maximum in FTIR-spectra close to the amyloid characteristic wave number of 1620 cm^{-1}. Therefore, it could be concluded that the CsgA generated by protein splicing folds into amyloid fibrils during the

reaction. In order to evaluate the isotope labeling efficiency, a ^{15}N-labeled sample was produced and measured in DMSO-d_6 by solution NMR. Figure 23 d shows a region of an overlay of 2D ^{1}H-^{15}N HSQC spectra of uniformly and segmentally (R1) labeled CsgA. A strongly decreased number of peaks were detected in the segmentally labeled sample while just a slight peak shift occurred in comparison with the fully labeled spectrum which occurred due to slightly different water content in the sample and small temperature differences during the measurement. By comparing the peaks positions to the sequence assignment it was confirmed that only the expected peaks were visible in the spectrum. Since the splicing product showed a clear amyloid character as well as the expected labeling pattern, all requirements to step on to solid state NMR were fulfilled.

For solid state NMR, ^{13}C/^{15}N labeled R1IntN fusion protein and unlabeled IntCR25 were used in a splicing reaction. The resulting pellet was washed several times with 4M GuHCl buffer and water and then spun into a 3.2mm MAS rotor. The final purity of the segmentally labeled CsgA sample for solid state NMR was estimated to 80-90% by SDS-PAGE. A two dimensional ^{13}C-^{13}C correlation spectra was recorded and compared to the spectrum of wild type CsgA. An overlay of the aliphatic region is depicted in figure 24 a. The peak positions were consistent with those of the uniformly labeled sample, but the spectral overlap was decreased. Taking into account that only spin systems involved in secondary structure elements might be detectable, according to the H/D exchange data, around 20 residues were expected for the R1 sample. The expected number of residues with a characteristic chemical shift like isoleucine, alanine, threonine and Serine could be unambiguously identified and the isotope labeling pattern confirmed. The alanine cross peaks are highlighted in figure 24 b, isoleucine cross peaks are shown in figure 24 c. The signal to noise ration appeared to be much lower than from uniformly labeled samples. Since both spectra were recorded on the same spectrometer the linewidth of the segmentally labeled sample was expected to be as narrow as the one of the uniformly labeled spectra. Even a minor improvement could be expected due to less labeled surrounding nuclei and therefore reduced destinations for spin diffusion. Unfortunately, the linewidth was much broader compared to the uniformly labeled sample and, due to the combination of low signal and broad lines, the resolution was lost for wide ranges of the spectrum. The low signal to noise ratio and broad lines are signs of an overall heterogeneity in the sample. Subsequently the generally weaker Nitrogen signal was not strong enough to record N-C correlations which are crucial for sequence specific assignments. The heterogeneity was most probably caused by contaminations and the not well defined fibrillization conditions during the splicing reaction. The Intein fusion proteins, containing parts or CsgA can readily influence the fibrillization of the ligated full length CsgA. Just little improvement was obtained by harsh washing of the sample so that finally the sample preparation strategy had to be changed in order to achieve more controlled fibrillization conditions resulting in more homogeneous samples.

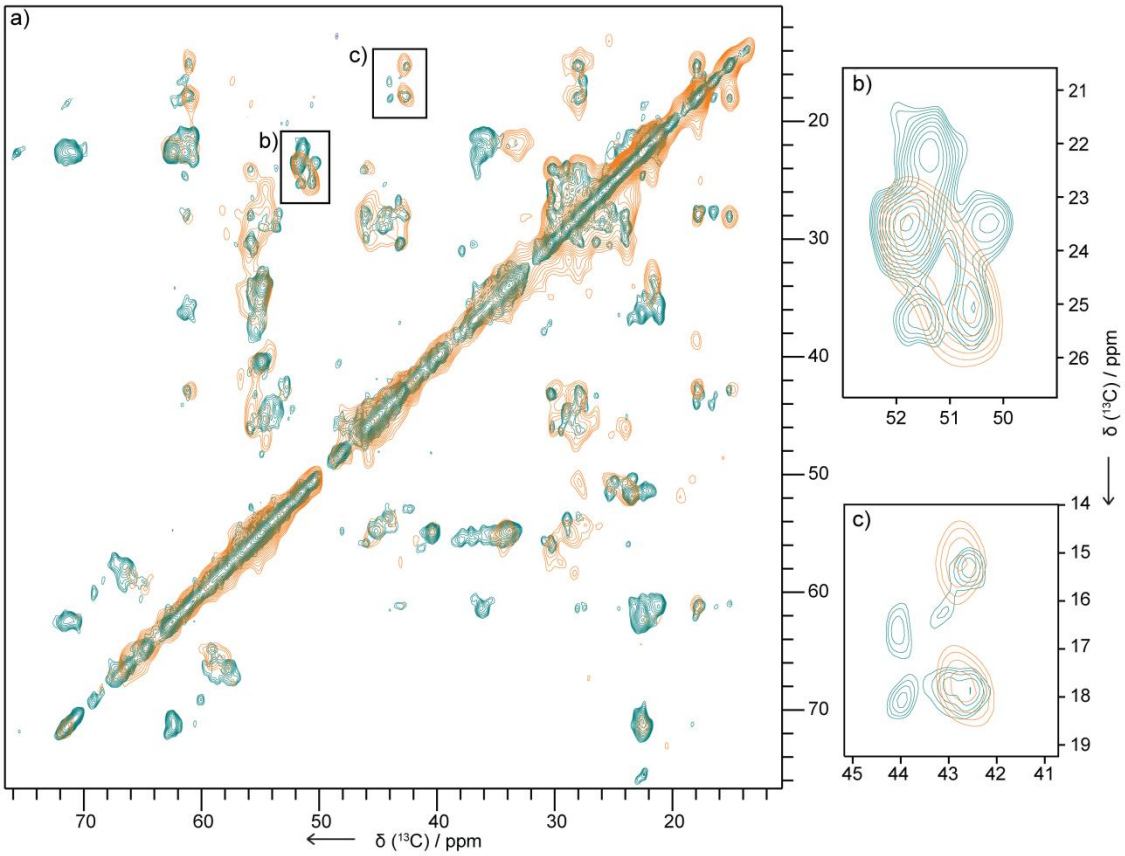


Figure 24: a) Overlay of the aliphatic region 2D DARR spectra of uniformly labeled CsgA (21-152) (teal) and segmentally labeled CsgA (^{13}C 21-70) (orange) measured at 21 T. b) Highlighted alanine Cα-Cβ cross peaks, c) Highlighted isoleucine Cβ-Cγ2, Cβ-Cδ1 cross peaks

4.1.5 Strategy Improvement

To improve the sample, pure and monomeric segmentally labeled CsgA had to be produced prior to fibrillization. Any further purification of the splicing product required either the prevention of aggregation during the reaction or the fibril disassembly after the reaction by dissolving in suitable buffers. The natural split intein *Npu* DnaE has been shown to be active in buffers containing up to 6M Urea. The visible protein aggregation during the splicing reaction of IntNR1 and IntCR25 could be prohibited in 5M Urea buffers but at the same time the splicing efficiency was reduced in an extent that a further elaboration of this strategy was not undertaken.

Amyloid fibrils of CsgA are highly stable and not soluble in commonly used denaturing buffers, 8M urea or 6M guanidine Hydrochloride. Formic acid (FA) and hexafluoro-2-propanol (HFIP) are commonly used to dissolve CsgA aggregates for SDS-PAGE analysis. For NMR samples, treatment with FA should be avoided since MALDI-MS revealed formylation of several residues in FA dissolved CsgA samples which could lead to complications or heterogeneity during the fibrillization. Additionally, for further purification steps both FA and HFIP need to be vaporized because they are not compatible with any biochemical chromatography material. Consequently neither FA nor HFIP is perfectly suitable for a continuous workflow using high protein quantities.

Buffers containing 5M guanidine thiocyanate, most commonly used for the extraction of DNA and RNA but also known to solubilize amyloids, were found to dissolve CsgA aggregates to high concentrations within a few minutes enabling further purification of the segmentally labeled samples.

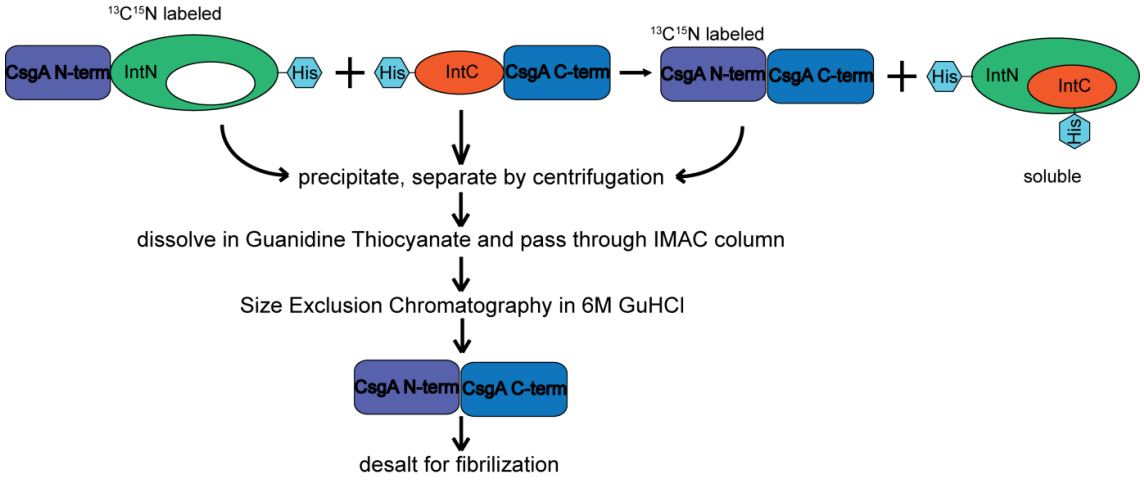


Figure 25: Optimized strategy for the production of segmentally labeled CsgA. The affinity Tags were bound to the intein domains so that the trans-splicing product is tag free. Precipitated proteins got dissolved in guanidine thiocyonate and CsgA further purified by passing though an affinity column (IMAC) and additional size exclusion chromatography. Pure monomeric CsgA was fibrilized by buffer exchange over a Sephadex desalting column.

Extensions at the C-terminus of the intein N domain have not been described before but based on the structure of the intein complex a C-terminal hexahistidine Tag should not interfere with the formation of the active state. The His-Tags of all three CsgA-IntN fusions were changed from the N- to the C-terminus, resulting in fusion constructs carrying tags only at the intein domains. The resulting splicing product would therefore be Tag-free and could be further purified by passages through an affinity chromatography column. By this step, all precipitated precursor proteins as well as leftovers of the intein complex were supposed to bind to the IMAC column while the tag-free CsgA passes trough. Additionally, a size exclusion step was added to the protocol to collect entirely monomeric CsgA and remove every kind of preformed aggregate and possible intein cleavage products. The improved sample preparation strategy is summarized in figure 25.

Preformed amyloids, so called seeds, can be used to improve the homogeneity of in vitro fibrillization. Usually amyloid forming proteins are kept monomeric under denaturing conditions and the fibrillization is initiated by buffer exchange to native conditions. By addition of Seeds right after desalting the fibrillization is accelerated and the monomers adopt a similar fold like the seeds. In case of CsgA, curli can be purified from *E.coli* biofilms and used to promote the native amyloid fold. To improve the homogeneity of segmentally labeled samples small amounts of purified native curli (provided by Puwei Yuan) were added to freshly desalted CsgA monomers.

4.1.6 CsgA R1 sample preparation (optimized)

The intein fusion R1IntN could be expressed in high yields from uniformly labeled minimal media while the relatively low expression yields of unlabeled IntCR25 were overcome with higher culture volumes. The position of the affinity tag did not show any influence in the expression yield. For protein trans-splicing, the purified precursors were mixed equimolarly and incubated at room temperature. The splicing progress could be monitored by continuously collecting samples for SDS-PAGE analysis.

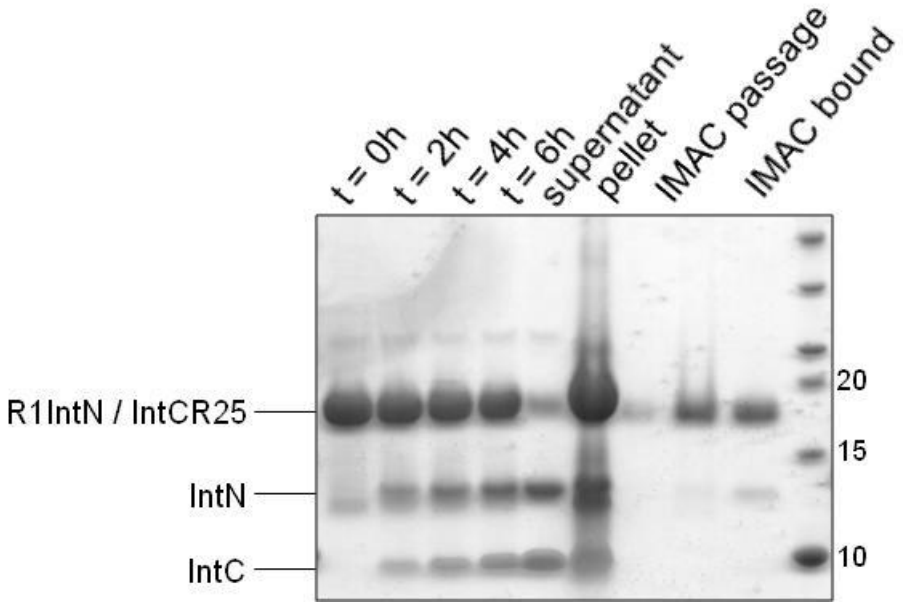


Figure 26: SDS-PAGE analysis of the R1IntN – IntCR25 splicing reaction monitored by the appearance of intein Bands after 2, 4 and 6 hours and product purification by reversed affinity chromatography. The precipitated educts as well as leftovers of the intein complex were successfully removed from the sample.

Figure 26 shows a SDS-PAGE analysis of R1IntN - IntCR25 splicing reaction and product purification by reversed affinity chromatography. The C-terminal His-Tag at the intein domain did not influence the splicing reaction: the intein domains appeared with equal intensities at comparable time points on the gel like with N-terminal His-Tag. Unfortunately, it was not possible to discriminate the educts (R1IntN & IntCR25) and the product (CsgA) from each other on SDS-PAGE gels, since they all ran on exactly the same molecular weight. However, it could be concluded that some of the educts stayed in the supernatant after centrifugation of the reaction mixture, what confirmed their solubility. The pellet showed a very strong band at around 18 kDa which corresponds to both precipitated educts and the product. The entire precipitated protein was dissolved in 5 M guanidine thiocyanate and the product was further purified by passage through an affinity chromatography column. The flow through as well as the bound protein were analyzed on the gel which clearly showed that some protein passed through the column, like expected for CsgA, and others of the same size bound the column, like

expected for precipitated educts. Additionally, the intein leftovers were completely removed from the sample solution.

Assuming a similar yield of the splicing reaction as found for the initial strategy a loss of target protein occurred during passages through Ni^{2+}-Sepharose column. This indicates an incomplete aggregate disassembly and accordingly a high affinity of the full length CsgA to the intein fusion proteins which results in intein mediated binding to the affinity column. Also the formation of disulfide bridges between educt and product molecules and ensuing binding to the column might decrease the purification efficiency. This effect can just partially be repressed due to the low compatibility of IMAC material with reducing agents. An additional complication is that the binding affinity of his-tagged proteins in 5M guanidine thiocyanate buffers is very little and several passages through affinity columns were necessary. No target protein could be detected on SDS-Gels after 5-6 passages because of unspecific binding. The efficiency of the reverse chromatography was improved by dissolving the sample to high concentrations (>20 mg/ml) and consequent dilution with 6M GuHCl buffer to reduce the thiocyanate concentration to < 2M.

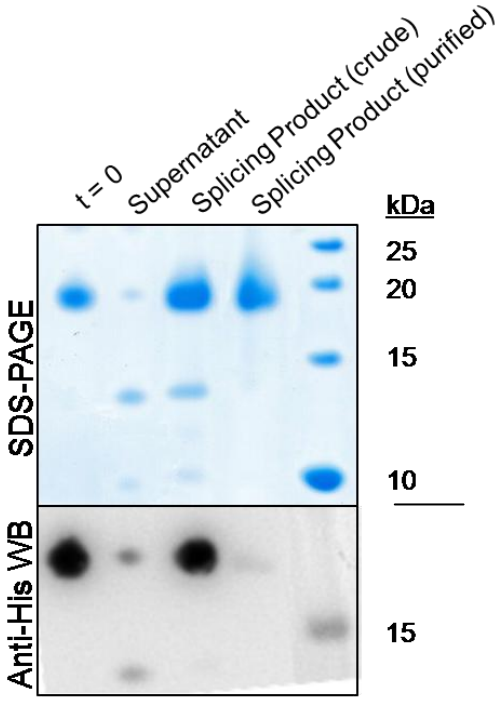


Figure 27: Analysis of the splicing product by SDS-PAGE and Western Blot. While both the crude and purified sample showed a strong band by Coomassie staining, western blot revealed that all His-Tagged contaminations could be removed from the sample and only the tag-free CsgA remained.

In a following experiment both the splicing and the purification products were analyzed in parallel by SDS-PAGE and anti His-Tag western blot (Figure 27). Like previously described, a strong band could be detected in SDS-PAGE at a molecular weight of about 18 kDa which corresponds either to the educts or to the product. By western blot just the crude sample without any further purification was detectable, the purified sample did not show any His-Tag signal, proving that the contaminations could be removed and exclusively the Tag-free CsgA remained.

In order to remove preformed aggregates and potential intein cleavage peptides which are not detectable on SDS-Gels, an additional size exclusion chromatography in 6M GuHCl was

performed. The chromatogram in figure 28 showed that aggregates of various sizes and a peak corresponding to a dimer (65ml) could be removed from the sample. The main peak at a retention volume of 78 ml corresponded to monomeric CsgA which was successfully separated from potential cleavage side products (retention volume of 115 ml). Additional peaks at a higher volume corresponded to DTT which was used as reductant and other buffer components (since this peak was much stronger, the chromatogram was cut at 117 ml for representative reasons). Afterwards, the protein solution was buffer exchanged from denaturing to native conditions to initialize the fibrillization and 2µl of a Curli suspension were added to initiate a controlled aggregation.

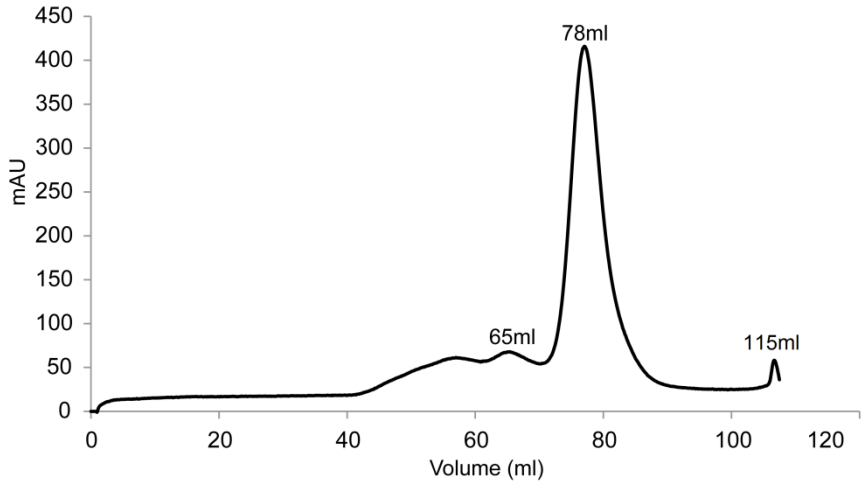


Figure 28: Size exclusion chromatogram of segmentally labeled CsgA (R1). After reversed affinity chromatography a Superdex 200 with 6M GuHCl running buffer was used to remove aggregates (65ml) and cleavage side products from the splicing reaction. Entirely monomeric CsgA (78ml) was collected for homogeneous fibrillization. The peak at 115 ml corresponded to cleavage side products with low molecular.

4.1.7 CsgA R2 sample (optimized)

The uniformly labeled intein fusion protein with C-terminal His-Tag containing the R1 and R2 repeats of CsgA was expressed in high yields (~35 mg/L culture). The protein trans-splicing reaction was not influenced by the affinity tag attached to the intein domain and occurred with the same kinetics like described before. The splicing product was collected by centrifugation and purified by several passages through affinity chromatography columns. The SDS-PAGE analysis of the splicing reaction and product purification by reversed affinity chromatography is shown in figure 29.

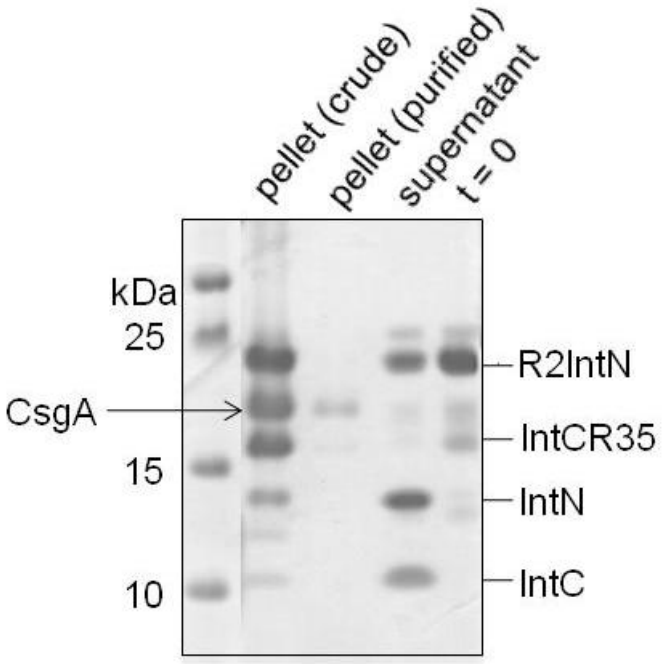


Figure 29: SDS-PAGE analysis of R2 sample production by protein trans-splicing of R12IntN and IntCR35. The intein domains were detected in the supernatant and CsgA in the pellet. Both educts precipitated during the reaction but were removed from the sample by reversed affinity chromatography.

Both intein domains were detected mainly in the supernatant, while the splicing product CsgA as well as the educts precipitated. All contaminations could be removed and entirely CsgA remained after purification by reversed affinity chromatography. The fading of the band was due to sample dilution during the purification process as well as sample loss due to unspecific binding of the product. Size exclusion chromatography was used as an additional purification step. A chromatogram of R1+2 labeled CsgA purification is shown in figure 30.

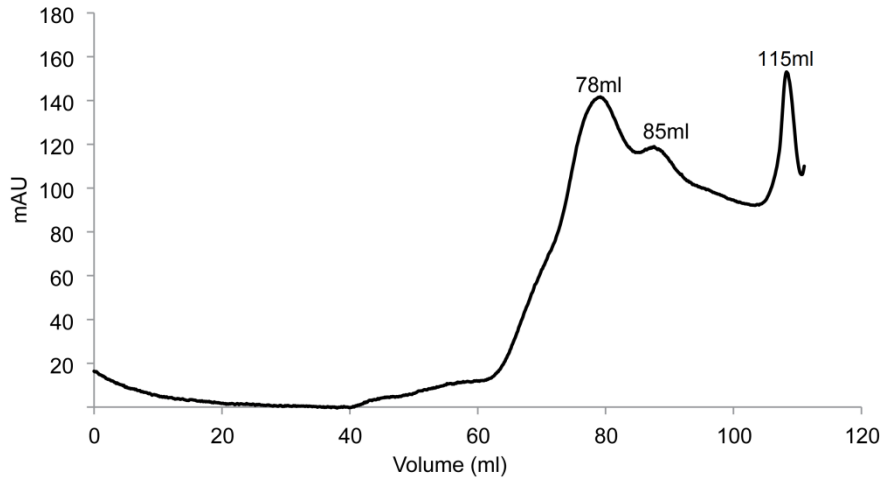


Figure 30: Size exclusion chromatogram of segmentally labeled CsgA (R1+2). After reversed affinity chromatography a Superdex 200 with 6M GuHCl running buffer was used to remove aggregates and cleavage side products from the splicing reaction. A strong contamination of R35IntC was detected at a retention volume of 85 ml. Fractions containing monomeric CsgA (78ml) were collected for homogeneous fibrillization; fractions close to 85 ml were avoided. The peak at 115 ml corresponded to low molecular cleavage side products.

The size exclusion chromatography revealed that large amounts of contaminations were still in the sample. The main peak at 78 ml corresponded to CsgA, while the second clear maximum, with almost the same intensity, corresponded to R35IntC. The peak at 115 ml, previously observed for the R1 sample and assigned as cleavage side products, was much stronger in the purification of R1+2 labeled CsgA but could be completely separated from the product. Also contaminations with higher molecular weight were found in the sample. In several trials to improve the reversed affinity chromatography step, the ratio of contamination was sometimes decreased but they were never completely avoided. For the solid state NMR sample preparation, just the middle peak fractions of CsgA were collected for desalting. The aggregation was seeded with wild type curli and folded into amyloid fibrils over night.

4.1.8 Towards a CsgA R5 Sample

In contrast to the R1 and R2 sample, for the R5 sample the inteinC fusion protein was supposed to be labeled with isotopes. Therefore, high and reproducible expression yields from minimal media were a requirement. In an initial unlabeled expression a yield of around 9 mg from one liter of culture could be obtained, which was a rough approximation since the quantification of this protein was complicated because it had a small extinction coefficient and was neither detectable on SDS-PAGE gels. A splicing reaction with R1-R4-IntN revealed the functionality of the fusion protein pair since the intein domains were detectable on a gel, as shown in Figure 31.

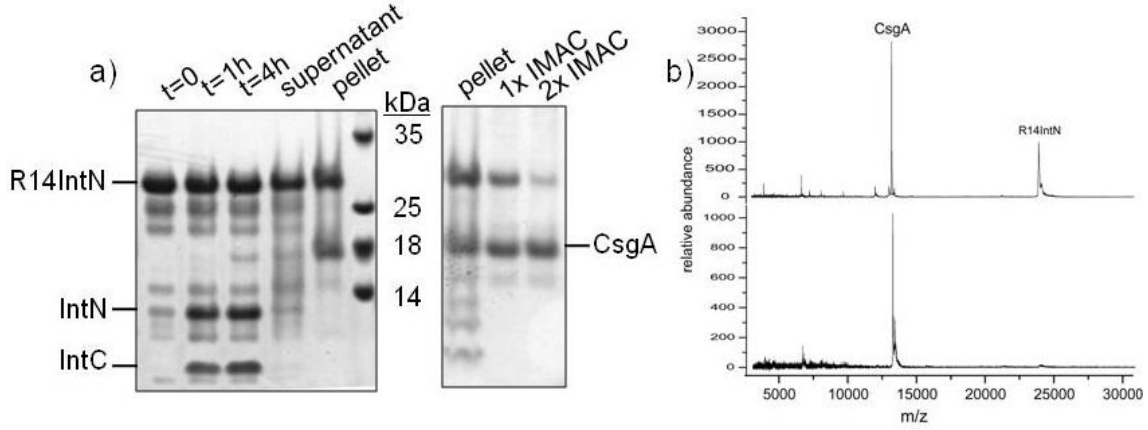


Figure 31: a) Splicing of R14IntN and IntCR5 (not visible) analyzed by SDS-PAGE. The intein domains appeared in the 1h and 4h time point samples. CsgA was detected in the protein pellet while R14IntN precipitated as well. The amount of R14IntN in the sample was significantly reduced after two passages through immobilized metal ion affinity chromatography columns. b) MALDI-MS analysis of the crude (top) and purified splicing products (bottom).

At the starting point only R14IntN and some artifacts caused by reaction with the sample buffer were visible on the gel. The splicing reaction occurred within several hours, comparable to the other reaction pairs. The trans-splicing product CsgA was found in the pellet, while R4IntN was detected partially in solution and in the precipitant. Reversed affinity chromatography revealed that several passages were necessary to completely remove R4IntN contaminations from the sample. Since IntCR5 was not detectable on a gel, the purification process was followed additionally by MALDI mass spectrometry, shown in figure 31 b. In the crude sample CsgA was detected with a molecular weight of 13110 g/mol as well as several contaminations like R4IntN

(23820 g/mol), IntC (3814 g/mol) and IntCR5 (7145 g/mol). Just CsgA was detected after purification as shown in figure 31 b. All contaminations could be removed from the sample by reversed affinity chromatography in 5 M GuTC buffer, while CsgA got purified and remained in the solution. This was particularly noticeable at the R14IntN peak.

Unfortunately these results were not straightforwardly reproducible due to very low expression yields of IntCR5 in minimal medium and quantification problems. Therefore several solubility Tags were tried in order to increase the expression yield and facilitate the quantification of the fusion protein. Additionally protein trans-splicing of IntC N-terminally fused with large proteins was an interesting subject of research because activity of those would permit segmental labeling of large insoluble, hydrophobic domains. The 55 kDa Nus-Tag, 22 kDa GST-Tag or the 11 kDa Trx-Tag were fused to IntCR5. Cleavage sites for either TEV or PreScission proteases were placed between the solubility tag and the IntC domain.

Nus-IntCR5 was expressed soluble and purified from the cytoplasm. The yields were high (>50mg/L) but the fusion protein was strongly degraded upon cell lysis. It was not possible to interrupt the degradation process by any protease inhibitor and therefore splicing reactions were not feasible.
GST-IntCR5 was expressed partially soluble but the purified protein appeared to be instable: it precipitated after reduction and no splicing product could be detected.
The biggest advancement was achieved with the Thioredoxin fusion. The majority of the protein was expressed in inclusion body and therefore the purification was undergone in denaturing conditions. The purified fusion protein could be refolded by dialysis and was stable for a limited time. Even with a loss of around 20% during refolding, the yields of functional protein were in an acceptable range for NMR (>30 mg/L). The stability was increased by dialysis against buffers containing 0,5-1 M urea.

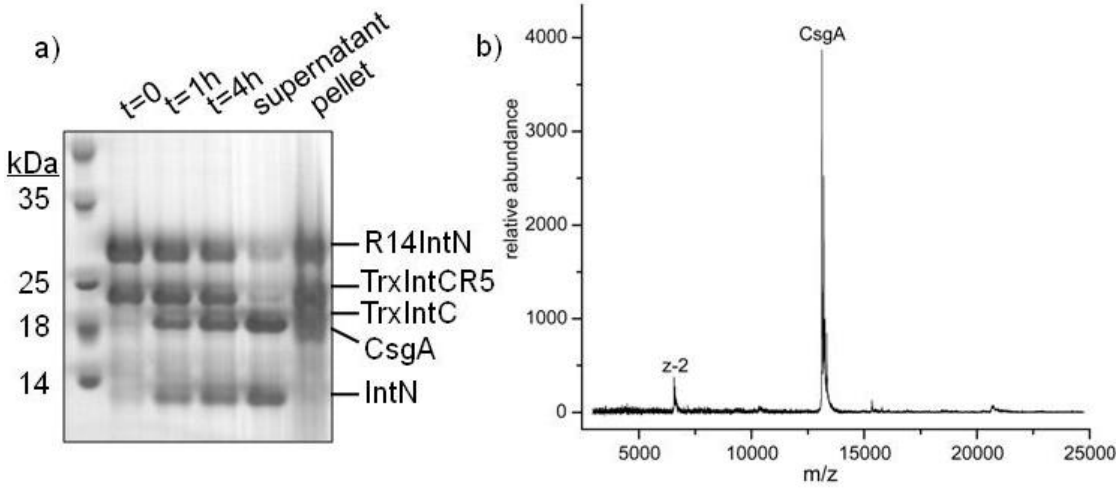


Figure 32: a) Splicing of R14IntN and TrxIntCR5 followed by SDS-PAGE. The detection of TrxIntC and IntN in the 1h and 4h sample indicated that the splicing reaction occurred. CsgA was detected in the protein pellet, also R14IntN and TrxIntCR5 precipitated during the reaction. b) MALDI-MS spectrum of CsgA produced by R14IntN-IntCR5 protein splicing, purified by reversed affinity chromatography

For the splicing reaction the precursor proteins (R4IntN; Trx-IntCR5) were mixed equimolarly in 0,5 M Urea buffer and the activity verified even without cleavage of the solubility tag as visualized on the SDS-PAGE gel in figure 32 a. The appearance of Trx-IntC and IntN served as indicators for the trans-splicing reaction. Both educts precipitated almost completely during the reaction, just a very faint band was detected in the supernatant after the reaction. CsgA could be detected in the protein pellet by MALDI mass spectrometry while the contaminations were removed by reversed affinity chromatography (Figure 32 b).

Further examinations like large scale splicing reactions and product purification by size exclusion chromatography still have to be performed. For solid state NMR investigations the focus of the present study was limited to the R1 and R2 samples.

4.2 Solid State NMR

4.2.1 R1 Sample

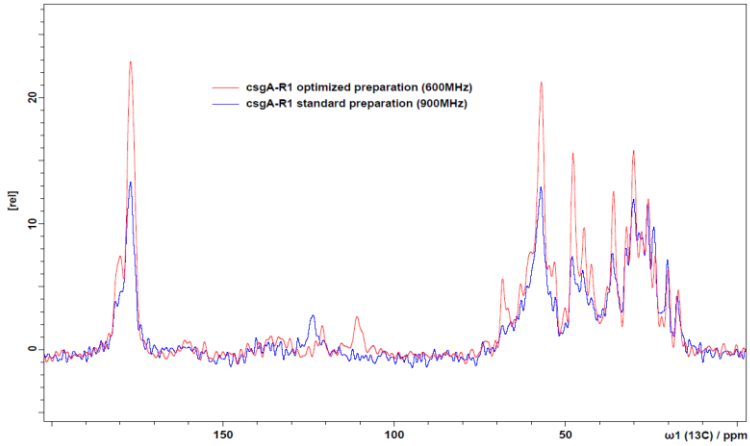


Figure 33: Changes in sample preparation strategy led to an improvement of signal to noise in 1D ^{13}C Cross polarization experiments.

To evaluate if the strategy changes led to an improvement of the sample in solid state NMR measurements, one and two dimensional ^{13}C spectra were recorded. Figure 33 shows an overlay of the 1D ^{13}C CP spectra recorded on the first sample and on the optimized one. The signal to noise ratio notably increased; estimated on the well separated CO peak (175ppm), the intensity was almost doubled. Beside the improved signal to noise ratio, also the peak linewidth was significantly reduced, as exemplified in figure 34 on the well separated methyl group cross peaks of isoleucine 47. All indications of heterogeneity like peak broadening and splitting were extinct and sharp isotropic lines were obtained. The peak positions of the ^{13}C-^{13}C cross peaks fit exactly on the ones of the uniformly labeled sample which again confirmed the high overlap which was strongly reduced using segmental isotope labeling.

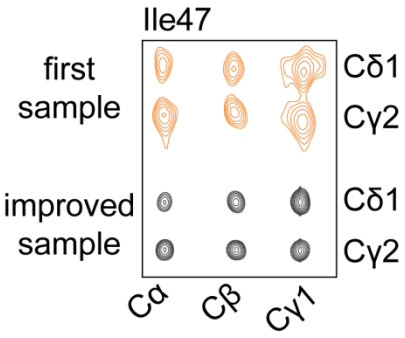


Figure 34: Changes in sample preparation strategy let to higher fibril homogeneity and therefore to an improvement of resolution and peak linewidth in 2D ssNMR spectra exemplified on the isoleucine 28 cross peaks.

The improved sample quality enabled the measurement of ^{15}N-^{13}C correlation spectra for sequence specific resonance assignments. First, resonances belonging to the same spin system were connected and, when possible, spin system types were identified using 2D ^{13}C-^{13}C, NCA and NCACX spectra, illustrated in figures 35, 36 and 37. The peaks are already labeled with the sequence specific assignment. Some unlabeled peaks in figure 35 correspond to assigned nuclei which are labeled elsewhere in the spectrum and some correspond to sequential distance constraints. Unlabeled peaks in figure 36 could not be assigned. The total set of spectra that were measured analyzed is listed in section 3.5.4.

The N-terminal 22 amino acids were not visible in the spectrum, probably due to flexibility. For example the isolated proline, which usually show a distinctive peak pattern, could not be detected. Since the CsgA R1 repeat contained just a single isoleucine and threonine, those residues could be identified and assigned by their characteristic chemical shift. Three alanines could be identified in the ^{13}C-^{13}C correlation and their nitrogen shift determined in the NCA spectrum. Also the expected three serine Cα-Cβ crosspeaks, the full set of leucine resonances as well as two asparagines could be identified. One arginine cross peak pattern could be identified and discriminated from leucine by the slightly different Cα chemical shift. Both glutamine residues overlapped in the ^{13}C-^{13}C spectrum but could be differentiated in the N-C. With the majority of spin systems identified it could be proceeded to sequence specific assignments.

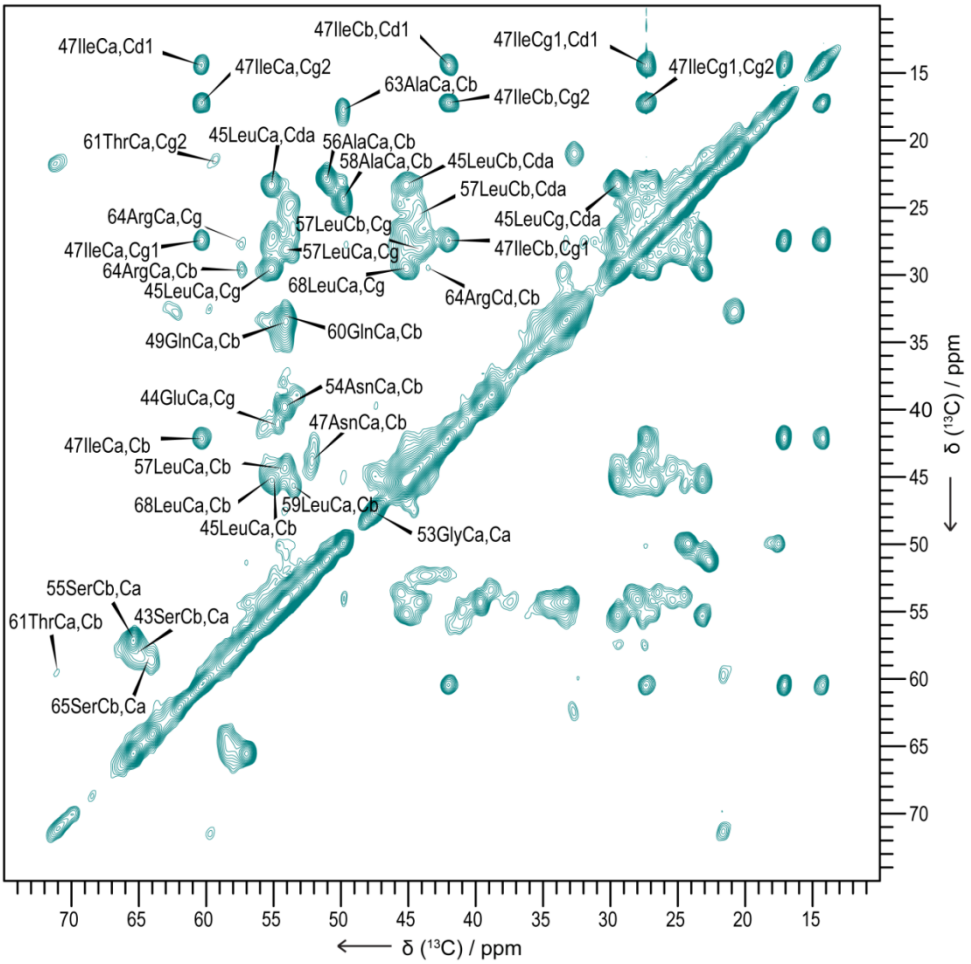


Figure 35: Aliphatic region of a 2D DARR spectrum acquired on segmentally labeled CsgA (^{13}C 41-90) at 21T with a mixing time of 50 ms. Cross peak assignments are marked on one side of the diagonal.

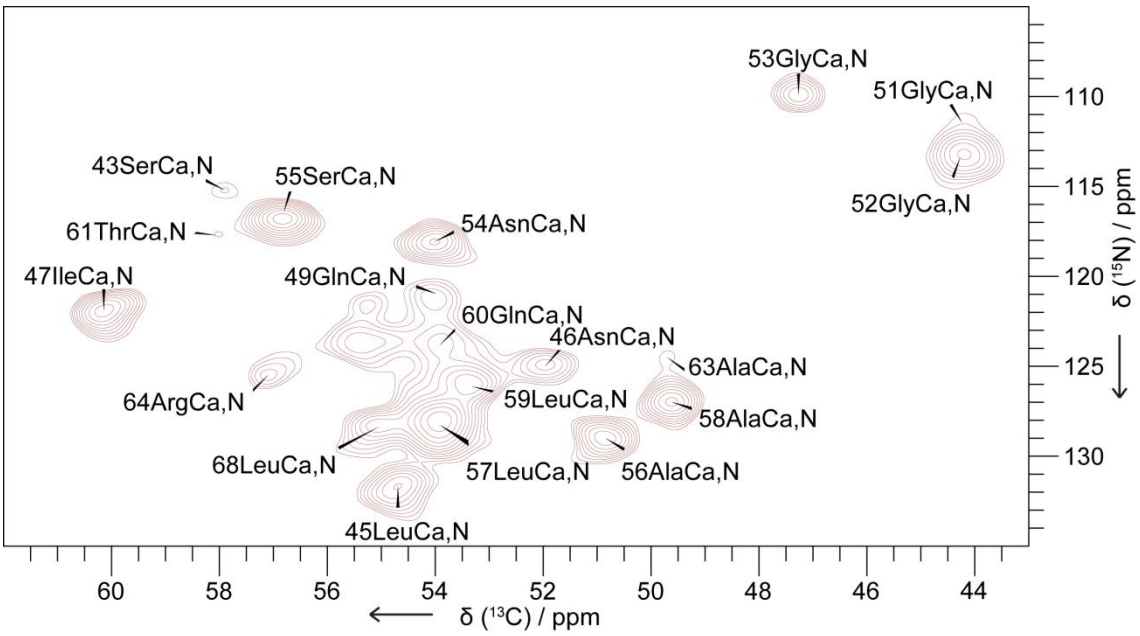


Figure 36: 2D NCA spectrum of R1 labeled CsgA amyloid fibrils recorded at 21 T. The sequence specific assigned peaks are labeled.

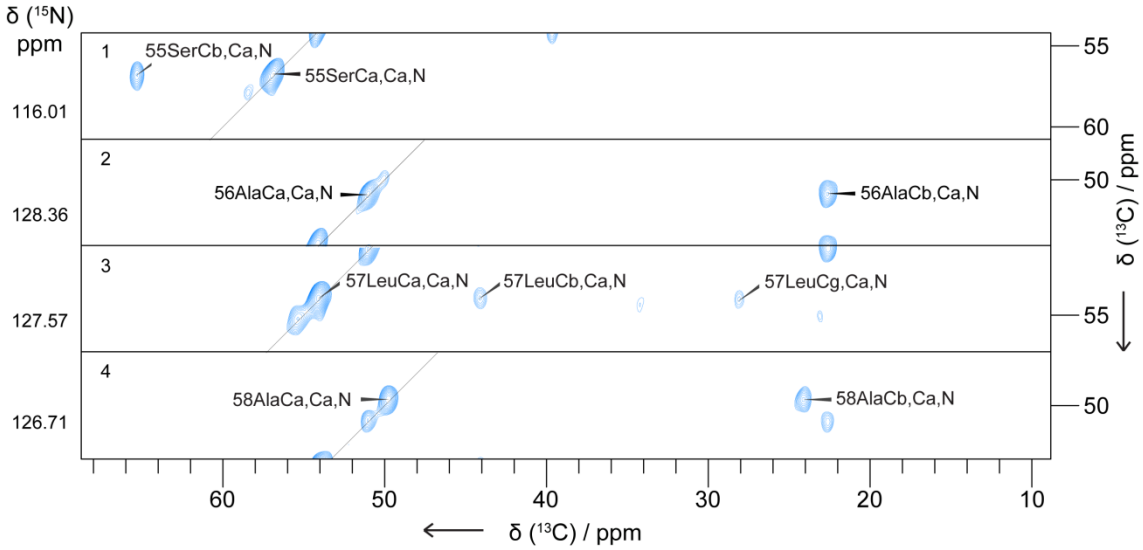


Figure 37: Strip plot of a 3D NCACX spectrum of R1 labeled CsgA fibrils recorded at 21T with a mixing time of 25 ms. The amino acid stretch Ser55, Ala56, Leu57, Ala58 is shown as representation. The nitrogen shifts (z-dimension in the spectrum) are annotated on the left. Residue specific assignments are labeled.

Sequential backbone connectivity was assigned using NCOCX and 2D ^{13}C-^{13}C correlation spectra with 300 ms mixing time (see figure 42). An exemplified backbone assignment of residues Gly53-Asn54-Ser55-Ala56 is shown in figure 38. By overlaying the NCOCX and NCACX spectra the Cα shifts could be sequentially assigned starting from the well separated alanine peak at ^{13}C 51 ppm and the next Serine peak at the same ^{15}N chemical shift. Since only a single A-S pair was present in the protein sequence, the spin systems could be unambiguously assigned. The preceding Asn and Gly residues were added exactly the same way, meaning by finding the preceding Cα at the actual nitrogen chemical shift.

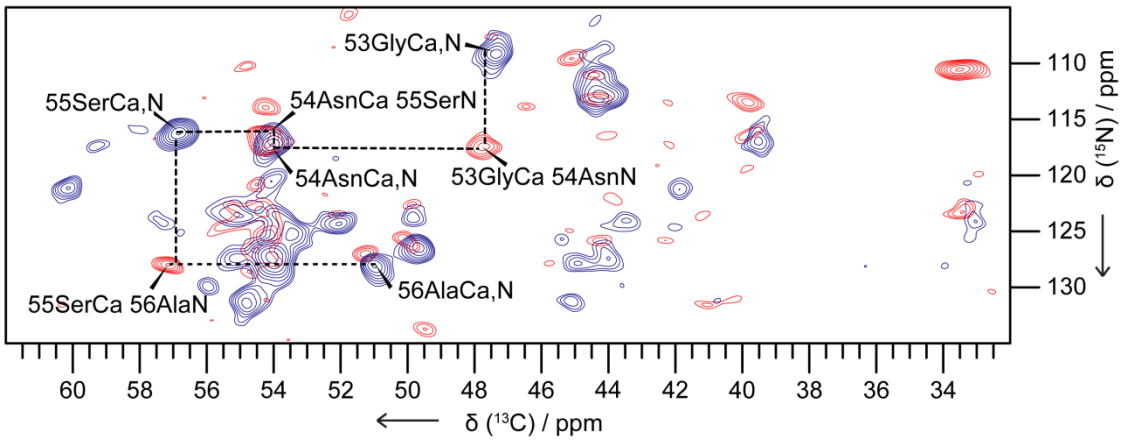


Figure 38: Overlay of 2D NCACX (blue) and NCOCX (red) spectra of R1 labeled CsgA amyloid fibrils. The sequential backbone assignment of 57Ala-56Ser-55Asn-54Gly is indicated by dashed lines starting from NCA of 57Ala, connected to SerCa at the same nitrogen shift in the NCOCX spectrum. This Ca can be identified at a different nitrogen shift in the NCACX and the sequential AsnCa at the same shift in the NCOCX. The backbone connection was continued the same way and let to a almost complete assignment of the R1 repeat of CsgA.

The sequential assignments were confirmed using a DARR spectrum, recorded with 300 ms mixing time, in which cross peaks correlations to neighboring spin systems were detected (figure 42). The side chain carbon chemical shifts were assigned either in NCACX or directly in ^{13}C-^{13}C correlation spectra. The rigid core of the segmentally labeled sample, from Ser44 to Thr61, as well as some residues from loop regions, could be assigned with strong evidence. Figure 39 gives a complete overview of the assigned resonances in the R1 repeat.

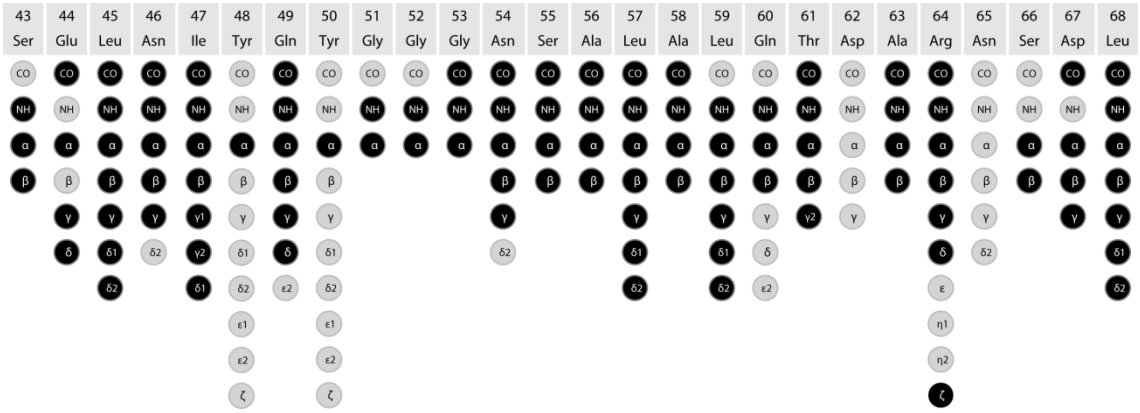


Figure 39: Assignment graph of backbone and side chain carbon and nitrogen resonances of the CsgA R1 repeat. Assigned resonances are depicted in black, unassigned ones in grey (plot generated with CCPN).

All cross peaks of Ser43 and Ser66 appeared much weaker than those of Ser 55, their CO´s could not be unambiguously assigned. Both tyrosines did not show any separated cross peaks and just the Ca could be identified over sequential restraints from either Ile47 or Gln49. Several intra-ring cross peaks were detected in the ^{13}C-^{13}C spectra but could not be clearly assigned to one of the residues. Asp62 and Asn 65 were not detectable and might be too flexible. No side chain nitrogen chemical shifts were assigned in general because none of the spectra provided sufficient information.

4.2.2 R2 Sample

The R2 sample was produced by protein trans-splicing of uniformly labeled R12IntN and unlabeled IntCR35. In comparison with the R1 sample, 20 additional residues were labeled with stable isotopes. A 1D ^{13}C CP experiment revealed that the signal to noise ratio was remarkably lower than the one of the R1 sample. Because of this, just a ^{13}C-^{13}C correlation spectrum was obtained, while the nitrogen signal was not intense enough to optimize for ^{15}N-^{13}C correlation spectra. An overlay of the R1 and R1+2 DARR spectra is shown in figure 40. Like expected, the majority of peaks appeared at the same positions due to the signal overlap. Additional peaks were evidently detectable and unique R2 amino acids could be assigned from the spectrum. The alanine 79 Cα-Cβ cross peak was clearly separated from the already assigned alanine peaks of R1. All isoleucine 71 side chain carbons were detectable at least at a single position of the spectra and could be added to the assignment. The characteristic chemical shift of valine methyl groups facilitated the assignment of valine 62 since it was the only valine in the CsgA R1 and R2 repeat.

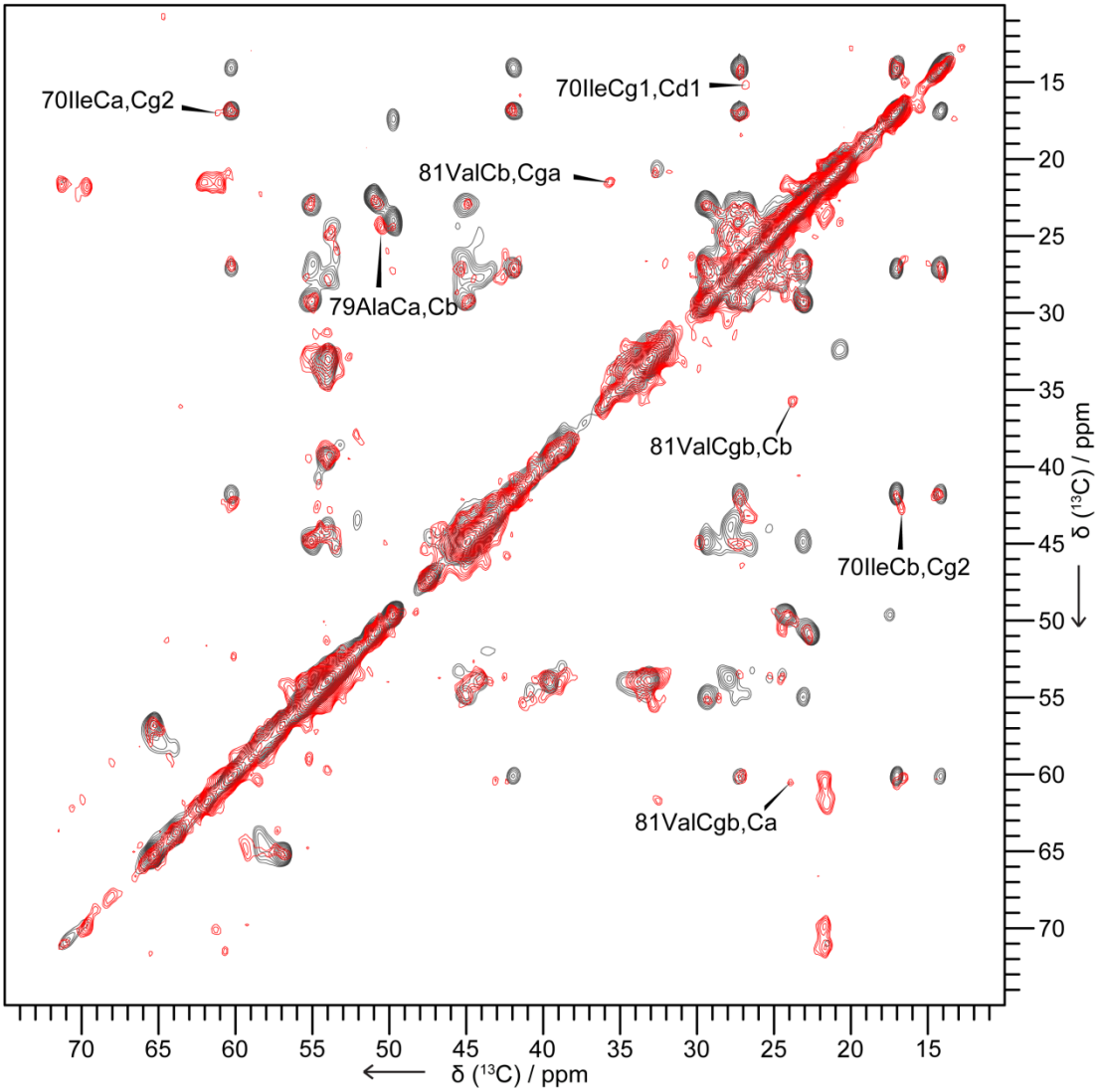


Figure 40: Overlay of the aliphatic region of 2D DARR spectra acquired on R1 segmentally labeled CsgA (^{13}C 21-70) (black) and R1+2 segmentally labeled CsgA (^{13}C 21-90) (red) at 21T with a mixing time of 50 ms. Unique residue chemical shifts gave rise to a partial assignment of R2 located residues which are marked in the spectrum.

4.2.3 Secondary structure from chemical shifts

The residue specific secondary structure elements were determined based on the Cα and Cβ chemical shift deviations from the random coil chemical shift. The primary sequence is listed on the top of Figure 37, where unassigned and assigned residues are lettered, respectively, in grey and in black. The chemical shift deviations of Cα and Cβ are depicted in figure 41 as filled bars, the predicted secondary structure elements are indicated by arrows under the sequence. The residues 21-42 were not detectable in the solid state NMR spectra and are most probably highly flexible and random coiled.

Two β-strands were identified in the R1 repeat, ranging from S43-Y50 and S55-T61, connected by a short glycin rich turn. leucine 69 was unambiguously assigned and might be the first residue of a following β-strand which includes also isoleucine 71 and ends by histine 73, assuming the same number of residues per β-strand in R1 and R2. The strands from R1 and R2 are connected by a 5-6 residue long loop. Moreover, the assigned A79, V81 and G82 were also involved in a β-strand, giving rise to the expected strand-loop-strand motive.

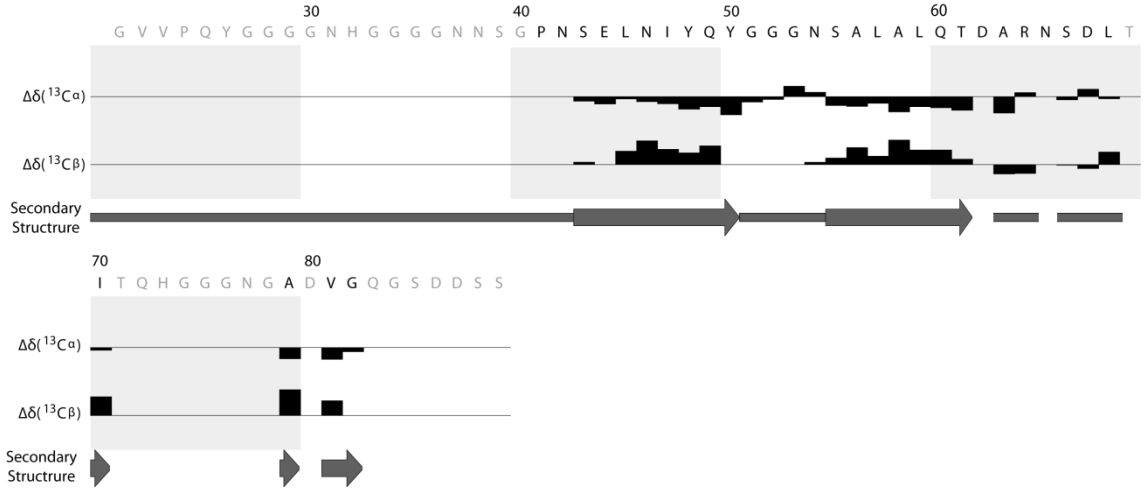


Figure 41: Sequence specific secondary structure. Chemical shift deviations of assigned Cα and Cβ from random coil shifts are indicated with filled bars. As secondary structure elements β– sheets are illustrated by arrows and random coil by straight lines.

4.2.4 Long range distance restraints

High resolution structure determination by solid state NMR relies on the identification of unambiguous sequential / short (+/- 2 residues), medium (3-4 residues distance) and especially long range distance restraints (≥ 5 residues distance). Usually ^{13}C-^{13}C correlation spectra like PDSD or DARR with long mixing times (up to 500ms) are exploited to detect long range restraints. Figure 42 shows a 300 ms DARR spectrum of R1 labeled CsgA amyloid fibrils. Several sequential links could be detected which confirmed the residue specific resonance assignment once more. Additionally, several medium and some long range distance restraints were identified especially for residues in the loop region between the two beta-strands (residues 50-56). The distance restraints unambiguously assigned to Ser55 and Asn54 are indicated in the spectrum and highlighted in figure 43 a. Both Ser and Asn show connections all around the loop including Gln49. This indicated a β-turn motive for this segment as illustrated in figure 43 b and gave rise to a strand-turn-strand fold in the R1 repeat of CsgA. Cross peaks arising from Tyrosine 50 were not detectable in the spectrum and the residue remained unassigned. No intersheet distance restraints could be identified from the 2D DARR, which represented the last crucial information for the calculation of a highly resolved structure of the CsgA R1 repeat.

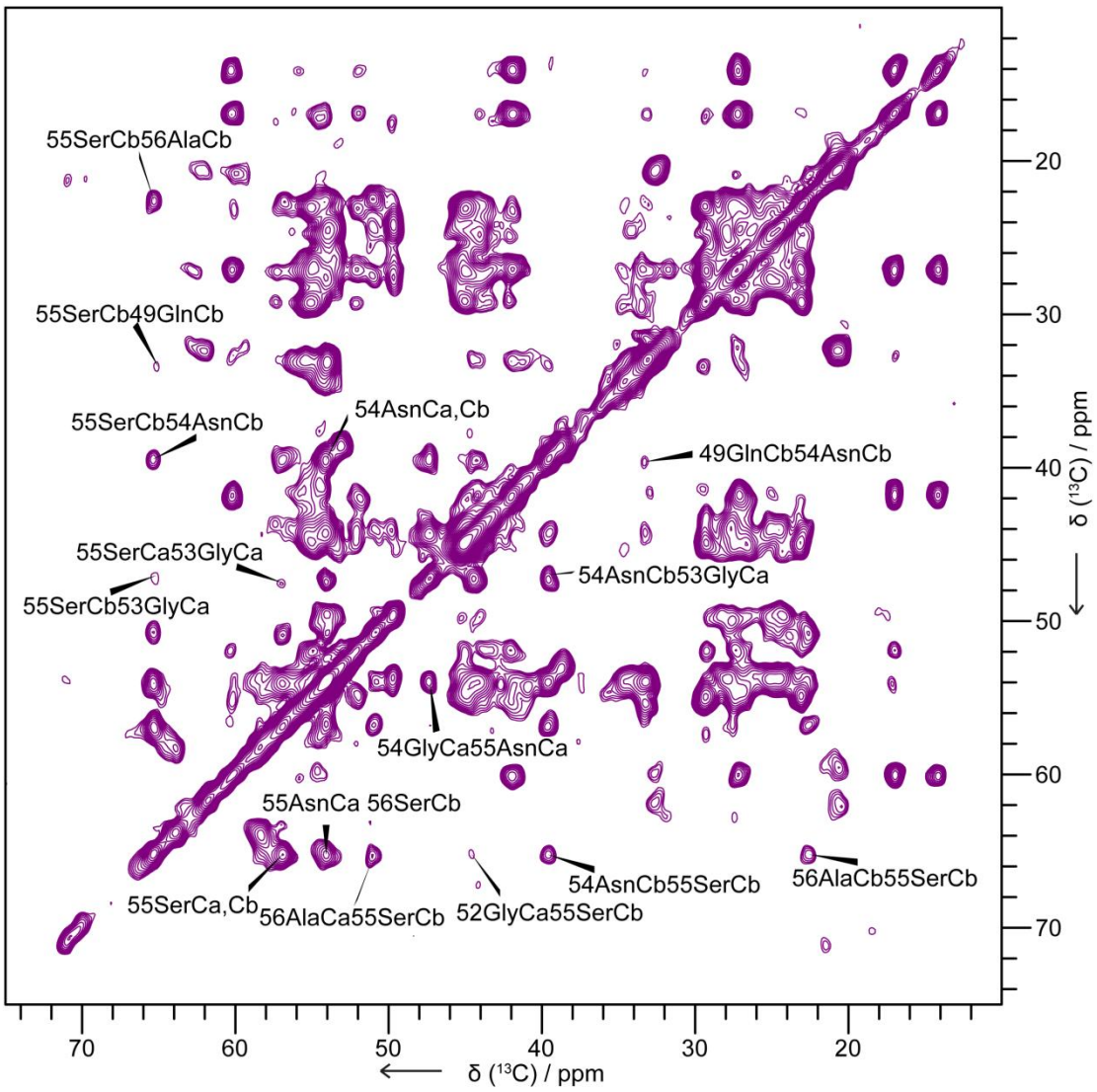


Figure 42: 2D DARR spectrum of R1 labeled CsgA fibrils, recorded with a mixing time of 300ms at 21T. Distance restraints detected for Ser56 and Asn55 are marked in the spectrum.

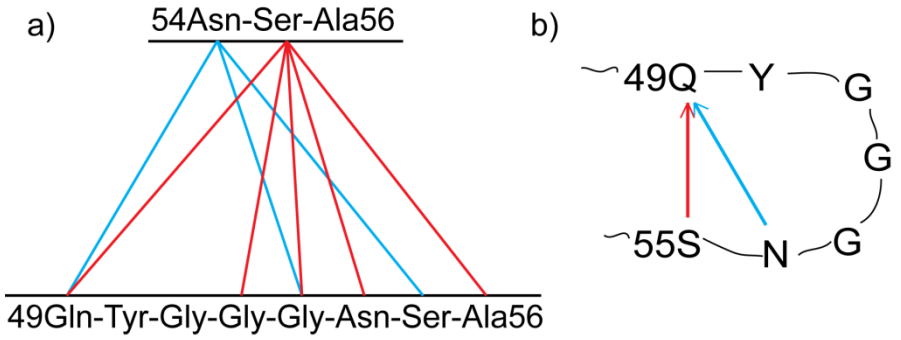


Figure 43: a) Graphical representation of distance restraints unambiguously assigned to Asn55 and Ser56. b) Interpretation of the assigned restraints from residue 50 to 55/56 giving rise to beta turn beta motive in the R1 repeat.

4.3 Segmental Isotope labeling of HET-s(218-289) amyloid fibrils

Through constant developments in magic-angle-spinning (MAS) solid-state NMR (ssNMR) methodology it became a powerful tool to determine structures of isotope labeled microcrystalline proteins, membrane proteins, and amyloid fibrils. Uniformly isotope labeled protein samples bear the highest amount of spectroscopic information. Complications like spectral crowding and peak overlap rise with the number of residues and the repeats within the amino acid sequence. Therefore, a combination of multidimensional NMR and isotope labeling techniques is commonly used for sequence specific assignments and the determination of distance restraints. Segmental isotope labeling is a promising approach to overcome spectral complexity in solid state NMR and to acquire highly resolved spectra of specifically labeled protein domains. The concept of segmental isotope labeling was adjusted for insoluble proteins shown for the bacterial amyloid adhesin CsgA. Here, the well characterized amyloid fibrils of HET-s (218-289) were used to prove the splicing and purification strategy on a system for which full chemical shift assignments and a structure are available. HET-s (218-289) molecules consist out of four β-sheets, forming a left handed β-solenoid.

Residue S257 was chosen as the splicing junction and substituted by the catalytic cysteine which is involved in the *trans*-splicing reaction. Fusion proteins with the intein N-terminal domain (IntN) and the C-terminal domain (IntC) have been generated with residues 218-256 and 258-289 respectively. In the HET-s amyloid, every monomer contributes two stacks of β-sheets to the fibril. In the resulting segmentally labeled sample, just a single layer would be labeled with isotopes, preventing the detection of intermolecular contacts by NMR. The expected labeling pattern is illustrated in figure 44 and the design of intein fusion proteins is illustrated in figure 45 a. IntN and IntC fusion proteins could be expressed and purified in high yields (~100mg/50mg per L culture) from inclusion bodies by denaturing with 6M guanidine hydrochloride. Refolding was initiated by dialysis and the proteins were stable as dimers in the presence of 1M guanidine hydrochloride. The same strategy, like described for CsgA, was followed for purification of the trans-splicing product. Histidine tags were connected to the intein domains so that the resulting product would be tag-free while all educts and co-products (intein complex) would carry an affinity tag. The strategy about how to obtain pure segmentally labeled Het-s fibrils is summarized in figure 45 b.

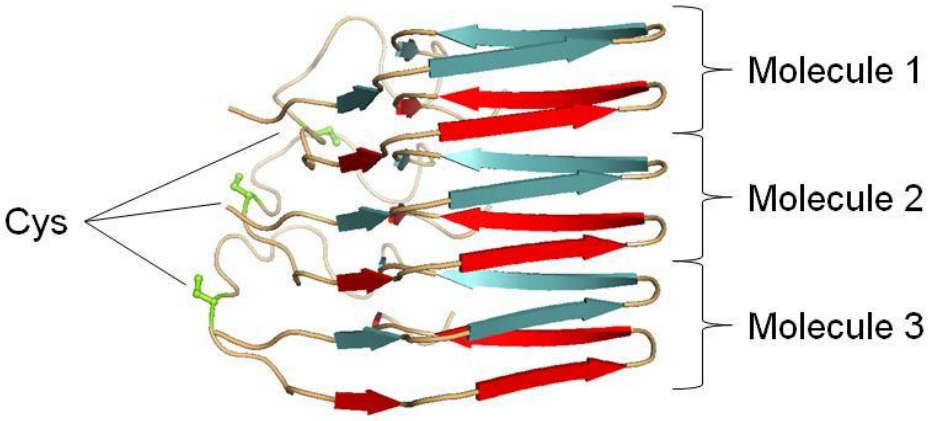


Figure 44: Layer specific segmental isotope labeling of HET-s (218-289) amyloid fibrils.

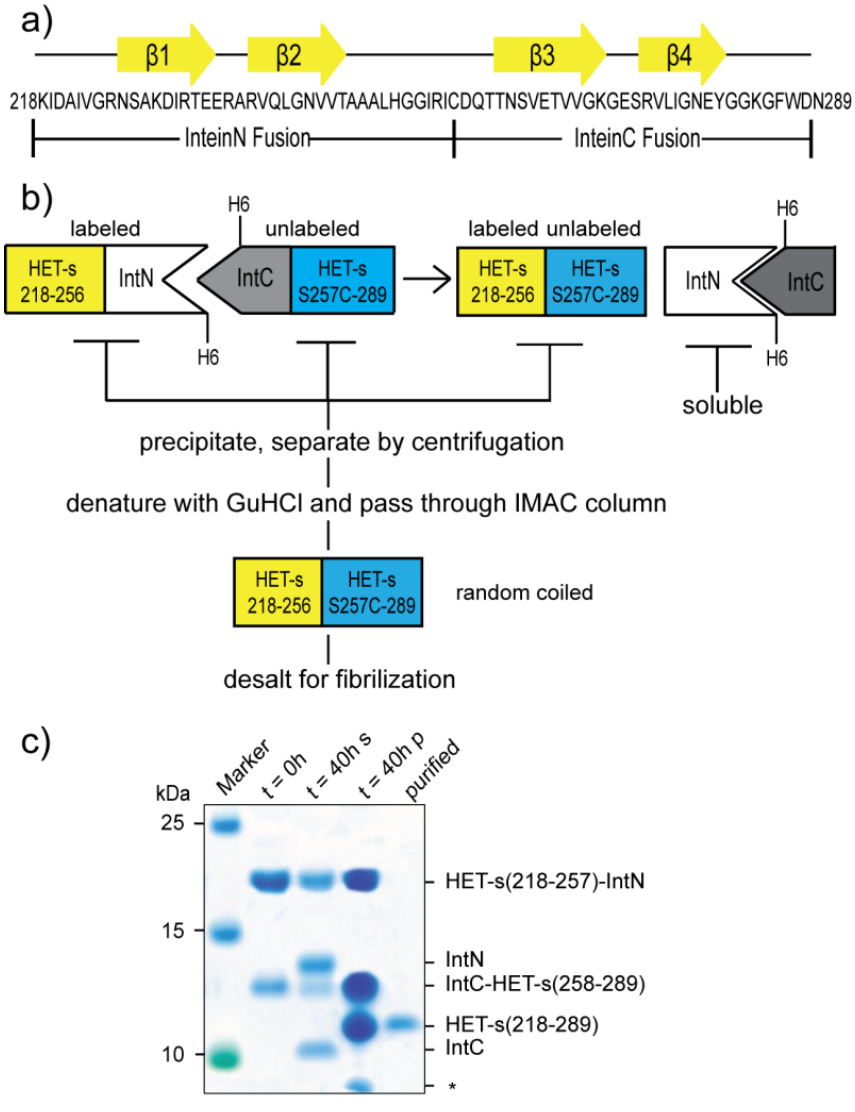


Figure 45: a) Primary and secondary structure of HET-s(218-289). The design of split intein fusion proteins is illustrated, S258 was chosen as the split side and replaced with the catalytic cysteine residue. b) His-tags were placed on the Intein domains; precipitated tag-free splicing product was purified from educts by reversed affinity chromatography in 6M GuHCl. The segmentally labeled HET-s folded into amyloid fibrils upon desalting. c) SDS-PAGE analysis of the splicing reaction of HET-s intein fusion proteins and the product purification by passages through a Ni^{2+}-Sepharose column.

The precursor proteins were mixed equimolarly and incubated at room temperature in the presence of TCEP as a reductant and 1M GuHCl. Under these conditions the intein complex stayed in solution while the trans-splicing product HET-s (218-289) precipitated. The two educts could be detected in the supernatant as well as in the pellet after 40 hours of reaction. The splicing efficiency was estimated to be around 20% based on the SDS-PAGE analysis (Figure 45 c). In contrast to CsgA, amyloid fibrils of HET-s are soluble in 6M GuHCl buffers which facilitated the product purification by reversed affinity chromatography. The precipitated educts as well as smaller side products that were detected in the crude pellet (*) could be removed from the sample by affinity chromatography. The purified segmentally labeled HET-s monomers folded spontaneously into amyloid fibrils after desalting into native buffer. Additionally, a uniformly labeled HET-s (218-289) sample was produced like described in chapter 3.3.4 in order to compare the spectra and investigate the advantages of segmental isotope labeling.

4.3.1 Solid state NMR of segmentally labeled HET-s (218-289)

To determine the labeling efficiency and the reduction of spectral crowding, 2D ^{13}C-^{13}C correlation spectra of uniformly labeled HET-s(218-289) and samples segmentally labeled either in the residues 218-256 or 257-289 were recorded. An overlay of the spectra is shown in figure 46, the cross peaks of uniformly labeled spectra are overlaid with only one of the segmentally labeled samples on each side of the diagonal, 218-256 above and 257-289 below the diagonal. The resonance lines of the spectra of segmentally labeled samples were narrow and comparable to those of HET-s(218-289) samples, indicating a highly ordered structure. The carbon backbone and sidechain chemical shift assignments were adopted from the Biological Magnetic Resonance Bank (BMRB), deposition number 11028. The Chemical shifts aberrations of the segmentally labeled samples to the published data and the uniformly labeled sample were less than 0.1 ppm, so that every resonance was unambiguously assigned and the labeling strategy confirmed. The spectral crowding was significantly reduced. Some regions, in which two or more peaks were not resolved in the uniformly labeled spectra due to chemical shift similarity, were clearly distinguished by the segmental labeling approach. Some examples are highlighted in figure 46 b, c and d.

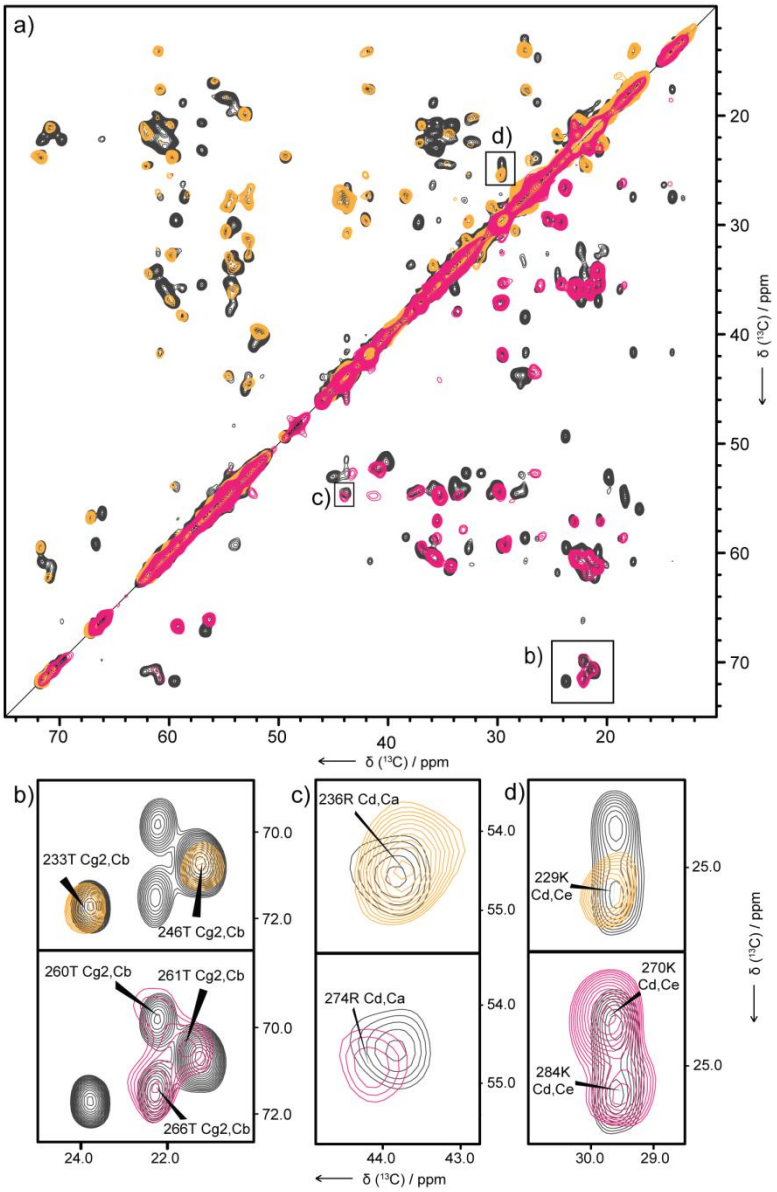


Figure 46: a) Overlay of the aliphatic region of 2D DARR spectra of uniformly labeled HET-s(218-289) (black), segmentally labeled HET-s (^{13}C 218-256) (orange) and (^{13}C 257-289) (pink) measured at 14T with a mixing time of 50 ms. Each of the segmentally labeled spectra are overlayed with the uniformly labeled one on just one side of the diagonal. b/c/d) highlighted regions of the spectra, demonstrating the high decreased signal overlap due to the labeling strategy

In order to evaluate the ability to measure unambiguous intramolecular distance restraints, ^{13}C-^{13}C correlation (DARR) spectra with a mixing time of 250 ms were recorded on on the HET-s (^{13}C 228-256) sample and the uniformly labeled sample. Extracted regions of the overlaid spectra are illustrated in figure 47.

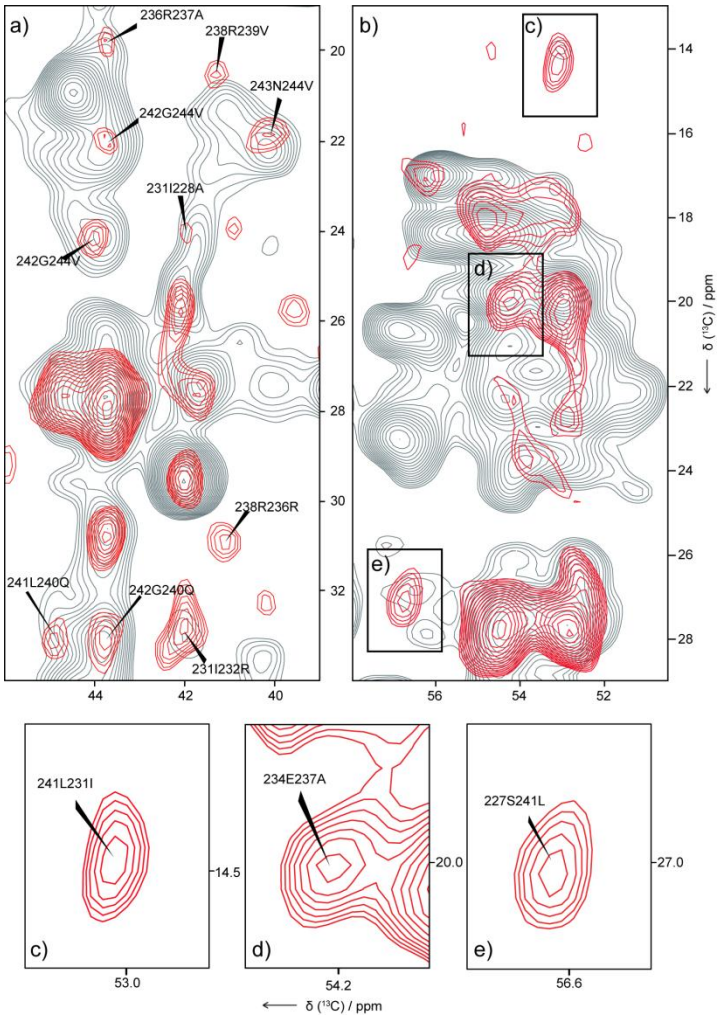


Figure 47: Overlay of the aliphatic region of 2D DARR spectra of uniformly labeled HET-s(218-289) (black) and segmentally labeled HET-s (^{13}C 218-256) (red) measured at 14T with a mixing time of 250 ms for the detection of long range distance restraints. Unambiguously assigned distance restraints are marked in the spectra. A region containing inter β-sheet distance restraints is shown in b) where three assigned peaks are additionally highlighted.

While in the spectrum of the uniformly labeled sample the great number of cross peaks gave rise to high ambiguity, the number of cross peaks in the segmentally labeled sample is decreased. Due to the labeling pattern which was alternated in the layers, intermolecular distance restraints could be avoided. Figure 47 a emphasizes the detection on sequential distance restraints which were much better defined in the segmentally labeled sample while mostly not resolved in the uniformly labeled sample. Figure 47 b shows a section of the spectra with focus on long range distance restraints, usually challenging to assign unambiguously. Several well separated peaks, three of them highlighted in further graphics, could be clearly assigned to intramolecular, intersheet contacts which are crucial for the definition of the tertiary structure (Figure 47 c,d,e). In total 53 unambiguous sequential and short (distance: 1-2 residues), 10 medium (distance: 3-4 residues) and 8 long range (distance: ≥5 residues) ^{13}C-^{13}C distance restraints were identified in the aliphatic region of the spectrum and listed in table 17.

Table 17: Distance restraints obtained from a 2D DARR spectra (250 ms) of HET-s(^{13}C: 218-257)

Sequential and Short Distance Restraints						
227	Ser	C	–	228	Ala	Ca
227	Ser	Ca	–	226	Asn	Ca
227	Ser	Ca	–	228	Ala	Ca
228	Ala	Cb	–	229	Lys	Cd
229	Lys	Ca	–	230	Asp	Cb
229	Lys	Cb	–	228	Ala	Ca
231	Ile	Ca	–	230	Asp	Cb
231	Ile	Cb	–	232	Arg	Cb
231	Ile	Cb	–	230	Asp	Ca
231	Ile	Cd1	–	233	Thr	Cb
232	Arg	Ca	–	231	Ile	Cg2
232	Arg	Ca	–	234	Glu	Cg
233	Thr	Ca	–	234	Glu	Ca
233	Thr	Ca	–	231	Ile	Cb
233	Thr	Cb	–	234	Glu	Ca
233	Thr	Cg2	–	234	Glu	Ca
233	Thr	Cg2	–	232	Arg	Cb
234	Glu	Ca	–	233	Thr	Cg2
234	Glu	Ca	–	235	Glu	Ca
235	Glu	Cb	–	237	Ala	Cb
235	Glu	Cg	–	236	Arg	Cg
236	Arg	Ca	–	235	Glu	Cg
236	Arg	Ca	–	234	Glu	Ca
236	Arg	Cb	–	237	Ala	Cb
236	Arg	Cd	–	237	Ala	Cb
236	Arg	Cd	–	235	Glu	Cg
237	Ala	Ca	–	235	Glu	Cg
237	Ala	Cb	–	236	Arg	Ca
238	Arg	Ca	–	239	Val	Cb
238	Arg	Cb	–	240	Gln	Cg
238	Arg	Cd	–	236	Arg	Cb
238	Arg	Cd	–	239	Val	Cga
239	Val	Ca	–	241	Leu	Cg
239	Val	Ca	–	237	Ala	Ca
239	Val	Ca	–	238	Arg	Cg
239	Val	Cb	–	237	Ala	Cb
239	Val	Cgb	–	237	Ala	Cb
240	Gln	Ca	–	241	Leu	Cb
241	Leu	Ca	–	239	Val	Cgb
241	Leu	Ca	–	239	Val	Cga
241	Leu	Ca	–	242	Gly	Ca
241	Leu	Cb	–	240	Gln	Cb
241	Leu	Cda	–	239	Val	Cga
241	Leu	Cdb	–	240	Gln	Cg
242	Gly	Ca	–	244	Val	Cgb
242	Gly	Ca	–	240	Gln	Cb
242	Gly	Ca	–	243	Asn	Cb
242	Gly	Ca	–	244	Val	Cga
243	Asn	Ca	–	244	Val	Ca
243	Asn	Cb	–	244	Val	Cga
243	Asn	Cb	–	245	Val	Cb
244	Val	Ca	–	243	Asn	Cb
248	Ala	Ca	–	247	Ala	Cb

Medium and Long Distance Restraints						
226	Asn	Ca	–	239	Val	Cb
227	Ser	Ca	–	241	Leu	Cda
228	Ala	Cb	–	239	Val	Cga
228	Ala	Cb	–	241	Leu	Cda
231	Ile	Cd1	–	239	Val	Cgb
231	Ile	Cb	–	228	Ala	Cb
231	Ile	Cg1	–	228	Ala	Ca
231	Ile	Cg1	–	228	Ala	Cb
232	Arg	Ca	–	239	Val	Cgb
232	Arg	Cb	–	239	Val	Cga
233	Thr	Cg2	–	237	Ala	Ca
234	Glu	Ca	–	237	Ala	Cb
238	Arg	Ca	–	233	Thr	Ca
239	Val	Cga	–	232	Arg	Cb
240	Gln	Ca	–	244	Val	Cgb
241	Leu	Ca	–	231	Ile	Cd1
241	Leu	Ca	–	231	Ile	Cg2
241	Leu	Cg	–	228	Ala	Cb

5 DISCUSSION

Amyloid formation is a hallmark of several neurodegenerative diseases like Alzheimer's or Creutzfeld and Jakobs disease where it is the consequence of protein missfolding. Additionally, there is a precedent for several proteins to form amyloids in which the large filamentous aggregates represent the native and functional fold. Such functional amyloids share all biophysical properties with disease related amyloids, like a characteristic X-ray diffraction pattern and the binding of the dye Congo Red. Amyloids have been found in a variety of organisms and the analysis of their amino acid sequence revealed that there is no common sequence motif that determines amyloid formation (Greenwald and Riek 2010).
The understanding of functions and mechanisms of amyloids as well as of differences between disease related and functional amyoids relies on the elucidation of highly resolved structures. Solid state NMR is capable to determine structures of proteins difficult to access by X-ray crystallography or solution NMR, what has already been successfully demonstrated for amyloid fibrils formed by the C-terminal domain of HET-s, residues 218-289 (Wasmer, Lange et al. 2008, Tycko and Wickner 2013).

Curli fibrils are cell adhesins expressed by enterobacteria like *Escherichia Coli*. They are involved in the formation of biofilms as well as in host cell attachment. Curli fibrils consist of a protein named CsgA which is capable to form amyloids *in vivo* as well *as in vitro* (Barnhart and Chapman 2006).
The subject of this work was the investigation towards the high resolution structure determination of CsgA by solid state NMR. There is very little knowledge about amyloid structures and the structural information of CsgA as a functional amyloid would greatly contribute to the understanding of amyloid formations and mechanisms. Better understanding of molecular primary sequence dependencies, e.g. roles of the CsgA gatekeeper residues could also be obtained (Wang, Zhou et al. 2010).
Some inhibitors of CsgA fibrillization have already been identified (Cegelski, Pinkner et al. 2009) and the structure of CsgA fibrils might provide the molecular basis of small molecule binding, giving access to a platform for structure based drug design of fibrillization inhibiting small molecules that act also on disease related amyloids. Those molecules might bind into the amyloid core and inhibit the formation of hydrogen bonds and salt bridges before maturation or in a way that disturbs the fibril propagation. Additionally, the structure of CsgA is a key step in researching the Curli biogenesis and the role of Curli in *E.coli* biofilm formation.
During this work, segmental isotope labeling was exploited for NMR assignments. The labeling technique was previously demonstrated for soluble proteins and now adapted for insoluble proteins, as exemplified on CsgA and a model amyloid fibril system HET-s (218-289).

5.1 CRITICAL ASPECTS IN THE CSGA SAMPLE PREPARATION FOR NMR

The most important requirement for good signal to noise ratios and well resolved ssNMR spectra is a highly homogeneous sample. In the case of amyloids the critical step is the *in vitro* fibrillization of isotope labeled protein monomers. Fibrillization occurs either by spontaneous folding of previously unfolded precursors or by induced conversion of natively folded proteins e.g. by stirring or sonication (Wang, Smith et al. 2007, Orrú and Caughey 2011). In many cases, especially for disease associated amyloids, this results in a high degree of heterogeneity or polymorphism in the sample which causes peak broadening and low resolved spectra (Heise, Hoyer et al. 2005, Petkova, Leapman et al. 2005, Debelouchina, Platt et al. 2010). Fibril heterogeneity is often not detectable by electron microscopy and therefore arises at molecular level.

The *in vitro* fibrillization of CsgA is very sensitive and can easily be disturbed by several factors. Lots of previous work was undertaken in our group and suitable conditions for sample preparation for H/D exchange analysis were already described (Zimmer 2011). The CsgA sample preparation for solid state NMR faced comparable challenges which are deeper discussed in the following.

A high purity and low protein concentration are necessary. The best results were obtained when a monomeric starting condition was guaranteed by filtering the protein solution through a 30 kDa cutoff filter or by size exclusion chromatography. Fast desalting using a column is favored over dialysis. A controlled seeding by adding preformed CsgA fibrils or Curli can also improve the sample homogeneity. By this, the aggregation lag-time is decreased and the fibrils adopt the same conformation like the seed.

Additional important aspects of the sample preparation for ssNMR are the water content within the sample and the filling into the ssNMR rotor. The NMR signal intensity is dependent of the amount of sample filled into the rotor. The highest signal to noise can be obtained with microcrystalline samples which are most densely packed. In the case of amyloid fibrils a good balance between sample compression and humidity has to be found. Dried or even lyophilized samples suffer from intensive peak broadening caused by disturbance of the structure stabilizing hydrogen bond network. Ultracentrifugation was used to reduce the water content in CsgA samples. Afterword they were transferred to the rotor with a table top centrifuge. Centrifugation forces up to 100000 x g were used for CsgA. At spinning speeds up to 200000 x g, like described for HET-s samples, CsgA samples became too dry and the spectral quality was reduced (Van Melckebeke, Wasmer et al. 2010). Devices for direct sample ultracentrifugation into ssNMR rotors were developed diminishing the risk of sample drying during the transfer but unfortunately nothing comparable was available during this study (Bockmann, Gardiennet et al. 2009).

5.2 IN VITRO FIBRILLIZED CSGA EXHIBITS HIGH RESOLUTION SOLID STATE NMR SPECTRA

Functional amyloids like CsgA consist of an evolutionary optimized primary sequence to form amyloid fibrils which are expected to be homogeneous as demonstrated for HET-s (218-289) (Siemer, Ritter et al. 2005). In an earlier solid state NMR study only poor spectral quality has been obtained due to the sample preparation protocol (Shewmaker, McGlinchey et al. 2009). Major drawbacks were the fibrillization initialization by dialysis and the lyophilization of the sample. In a previous project conducted in our group, monophasic H/D-exchange rates indicating highly homogeneous CsgA samples were obtained (Zimmer 2011). More specifically, the H/D-exchange analysis revealed a beta-loop-beta structure comprising around 90 rigid residues.

Taking all aspects described in chapter 5.1 into consideration, reproducible homogeneous CsgA samples could be produced which resulted in solid state NMR spectra with high signal to noise ratio and narrow peaks. All 3 isoleucine and at least 6 alanine residues which are widely spread over the amino acid sequence could be identified, demonstrating a proper fold of the entire molecule. No indications about different isoforms of CsgA fibrils arose from the ssNMR data. The overall sharp linewidth of less than 1 ppm was confirmed by a segmentally labeled sample which comprised more isolated peaks. This was comparable to spectra of HET-s (218-289) amyloid fibrils and confirmed the assumption that the CsgA sequence is evolved to adopt a specific molecular structure without the polymorphism observed for Aβ1-40 or yeast prion proteins (Siemer, Ritter et al. 2005, Fandrich, Meinhardt et al. 2009) .

Three dimensional ssNMR spectra allowed the almost complete backbone assignment of proteins with more than 300 residues (Habenstein, Wasmer et al. 2011). Decently resolved 3D solid state NMR spectra could be obtained from uniformly labeled CsgA samples (data not shown) but due to the unfortunate cross peak distribution the backbone assignment based on ^{15}N-^{13}C correlation spectra remained challenging. Further developments in dynamic nuclear polarization ssNMR or ultra fast magic angle spinning in combination with proton detection might provide the missing resolution to obtain structural information of full length CsgA.

5.3 ISOTOPE LABELING STRATEGY SELECTION FOR THE ASSIGNMENT OF HIGHLY OVERLAPPED NMR SPECTRA

Uniformly labeled samples contain the highest amount of spectroscopic information but the spectra tend to be complex due to signal overlap. Therefore, isotope labeling strategies are exploited to reduce the complexity by decreasing the number of labeled nuclei in the sample. The two dimensional solid state NMR spectra of uniformly labeled CsgA fibrils suffer from an overall signal overlap (see Figure 18) caused by a highly repetitive amino acid sequence and structure. Therefore an isotope labeling strategy to reduce the spectral overlap and facilitate the residue assignment needs to be applied.

Several isotopic labeling strategies which have been used for backbone assignments and the detection of distance restraints in solid state NMR are available (Renault, Cukkemane et al. 2010). Each technique comprises advantages for specific applications but most of them would not lead to advancement in the specific case of CsgA.

The (forward) labeling of just one or a few types of amino acids is very expensive and would not contain sufficient information for sequence specific assignments because most of the sequential connections would be lost. Reverse labeling, meaning unlabeling of certain amino acids is generally used to assign cross peaks that are covered by peaks arising from a different residue type. In case of CsgA almost every type of amino acid faces this situation, except of the naturally well separated residue peaks of Ile, Thr, Ser and Ala. The resolution of the highly repetitive residue stretches would not improve.

The utilization of a checkerboard labeling pattern by sparsely labeled carbon sources would reduce the spectral crowding and decrease the linewidth by the removed J-couplings. The labeling technique can be exploited for assignments as well as the detection of long range restraints and has successfully been demonstrated e.g. for HET-s (218-289) amyloid fibrils (Wasmer, Lange et al. 2008). The reduction of overlap and improvement in resolution might not be sufficient for the sequence and structure repeats of CsgA.

Just segmental isotope labeling, meaning labeling just a defined part of a protein with isotopes, reduces the spectral complexity in such extend that sequence specific assignments become feasible (Volkmann and Iwai 2010). Furthermore, with a decent set of different samples, the strategy barriers the potential to provide all information necessary to calculate a structure of CsgA.

Protein splicing relies on pairs of fusion proteins each containing a part of the target proteins giving rise to several combinatorial possibilities which address different spectroscopic information. The number of labeled residues can be continuously reduces to obtain sufficient resolution for sequence assignment. For the fivefold repetitive sequence of CsgA, constructs that contain 1 to 2,5 repeats from the N- and C-terminal side can be generated to achieve stepwise a complete resonance assignment. In this case one part of the protein sequence remains unlabeled. Considering the continuously stacked β-sheets in an amyloid fibril, intermolecular distance restraints are eliminated and the detection of unambiguous intramolecular restraints facilitated as demonstrated for HET-s (218-289). Additionally several labeling combinations can be used to detect unambiguous intermolecular restraints such as ^{13}C labeling in one segment and ^{15}N or amino acid specific labeling in the other. Also a completive approach using 1- ^{13}C glucose and 2-^{13}C glucose could be exploited.

5.4 THE NATURAL SPLIT INTEIN *NPU* DNAE ENABLES TRACELESS SEGMENTAL ISOTOPE LABELING OF INSOLUBLE PROTEINS

Segmental isotope labeling is a promising technique to facilitate the analysis of multi domain proteins by nuclear magnetic resonance in solution or in the solid state. Protein trans-splicing is the most suitable approach for the labeling of proteins due to robustness, high ligation efficiency

and yields. It requires the expression of split intein fusion proteins, each containing one part of the protein of interest, followed by their ligation *in vitro* or *in vivo* (Muona, Aranko et al. 2010). Ideally, inteins would seamlessly splice polypeptides together with no remnant sequences to guaranty native characteristics and fold of the target protein (Lockless and Muir 2009). This is especially important for the segmental labeling of amyloid fibrils which tend to fibrillize heterogeneously. The *Npu* DnaE intein showed a high amino acid tolerance in a mutagenesis study of the Phe +2 position enabling a native-like construct design (Iwai, Zuger et al. 2006).

The natural extein sequence KCFNG pledges the highest splicing activity and has been introduced to the target protein sequence in previous segmental isotope labeling studies (Buchinger, Aachmann et al. 2010).

Here, an almost traceless protein splicing was achieved to produce segmentally labeled CsgA and HET-s(218-289) amyloid fibrils. The key residue in the catalytic mechanism is a cysteine between the intein c-terminal domain and the extein which was introduced and remained as single point mutation in the splicing product. It has already been demonstrated for CsgA and HET-s that single cysteine mutations are accepted and do not affect the fibrillization (Ritter, Maddelein et al. 2005, Zimmer 2011). The resulting extein sequences were: CITQH, CIDLT and CNSSV for CsgA and CDQTT for HET-s (218-289). In the mutagenesis study by Iwai and coworkers the activity was influenced by less than 10% for an Ile or Asn residue at the +2 position. An Asp residue at the +2 position reduced the activity to around 50% (Iwai, Zuger et al. 2006).

All intein fusion constructs could be expressed and purified in sufficient yields for NMR. All split intein pairs showed activity and proved the high tolerance of the *Npu* DnaE intein concerning the flanking extein sequence. No significant differences between the three CsgA pairs could be identified. The splicing efficiency of the HET-s intein fusion proteins turned out to be very low due to insolubility of the precursor proteins which could be improved by the addition of 2M Urea or 1M GuHCl during the reaction. The kinetics of the splicing reactions were not further analyzed since it is not of the highest relevance for the production of NMR samples. The critical parameter is the yield of labeled protein per liter of culture. Just a half liter of labeled media was necessary for the production of a R1 labeled CsgA which is less than the volume of media used for the production of a uniformly labeled sample. Therefore it was successfully demonstrated that protein *trans* splicing can be applied for segmental isotope labeling without a severe alternation of the target protein sequence and without the requirement of increased culture volumes.

Considering the highly hydrophobic primary sequence of CsgA it can be assumed that the same strategy can be realized with various other difficult targets like transmembrane helices. Here, an almost traceless ligation with just a single cysteine point mutation remaining was achieved. A completely traceless segmental isotope labeling becomes feasible if a native cysteine residue is available as a potential split side.

5.5 STRAIGHTFORWARD TRANS-SPLICING AND PRODUCT PURIFICATION FOR SEGMENTAL ISOTOPE LABELING OF SOLUBLE AND INSOLUBLE TARGET PROTEINS

Segmental isotope labeling by protein trans-splicing relies on an auto catalyzed ligation of polypeptides by split inteins. For soluble proteins and solution NMR studies high yields of segmentally labeled proteins can be obtained by splicing *in vivo* or *in vitro* (Muona, Aranko et al. 2010). The most important protein targets for solid state NMR are those which are not easily accessible by solution NMR or X-ray crystallography namely amyloid fibrils and membrane proteins in their native lipid environment. Those types of proteins usually consist of a highly hydrophobic and repetitive primary structure and, in *E.coli*, get expressed into inclusion bodies. Therefore the splicing reaction is just feasible *in vitro* after purification. Here, a straightforward strategy of fusion protein purification from inclusion bodies, protein splicing and product purification has been established.

Due to the inclusion body expression of the CsgA / HET-s intein precursors the purification requires denaturing conditions. For that reason the intein domains need to refold into the active state. It was possible to show that the Inteins domains can be easily refolded and are already active after dialysis from strong denaturing conditions (6M GuHCl) into native buffers, or, in case of low stability of the fusion proteins, into buffers containing 2-3 M urea which minimizes the loss of precursor protein. This allows protein trans-splicing of poorly soluble precursor proteins and enables segmental isotope labeling of difficult targets, like demonstrated for amyloid fibrils.

In an ideal scenario the two intein fusion proteins react completely into the ligated target protein and the intein complex. In reality the splicing reaction *in vitro* is not complete, meaning some educts remain and additionally intein cleavage side products occur in variable extends. The splicing product needs to get purified for NMR studies. For soluble proteins the entire compilation of affinity tags and purification methods are applicable while, for insoluble proteins, which have to be kept under denaturing conditions, the methods are rather limited. Affinity tags that rely on a proper fold of the peptide like Strep tag are not functional under denaturing conditions. Chromatography methods like ion exchange, commonly used to purify soluble splicing products, suffer from reduced separation resolution under denaturing conditions and are challenging to apply on proteins that consist in majority of the same primary sequence.

In this work, the intein fusion plasmids were designed to produce a tag free splicing product to permit a removal of precipitated educts by affinity chromatography. The His-Tags on both of the intein domains did not influence the splicing affectivity. Size exclusion chromatography is applicable under denaturing conditions and was used to remove smaller byproducts from the sample. With this strategy a sample purity of more than 90% was achieved. The protocol is transferable to every protein of interest which implies an elaborated refolding protocol. Together with the possibility of traceless protein splicing due to highly tolerant inteins, a step towards a more straightforward and general use of segmental isotope labeling for structural biology by NMR has been achieved.

5.6 Resonance assignments of highly overlapped solid state NMR spectra of amyloids by segmental isotope labeling

The feasibility of sequential resonance assignments on segmentally labeled CsgA samples was demonstrated. CsgA consists of 152 amino acids in total and the hydrophobic core region is build up of a fivefold imperfect repetitive unit (R1-R5). Additionally the same technique was applied to the well characterized amyloid fibrils of HET-s (218-289) confirming the labeling pattern on a fully assigned system.
Segmental isotope labeling requires the ligation of a labeled and an unlabeled part of a protein. This can be achieved by protein trans-splicing of split intein fusion proteins. The initial purification strategy of the splicing product was based on different solubility of the product and the educts. The resulting samples formed amyloid as demonstrated by FTIR-spectroscopy and electron microscopy but suffered from a severe microheterogeneity, revealed by solid state NMR. Subsequently the generally weaker Nitrogen signal was not strong enough to record N-C correlations which are crucial for sequence specific assignments. The heterogeneity is most probably caused by contaminations and the not well defined fibrillization conditions during the splicing reaction. The intein fusion proteins, containing parts or CsgA can readily influence the fibrillization of the ligated full length CsgA. A less likely but possible cause of heterogeneity is the cysteine point mutation arising from the splicing reaction.

Like described earlier for the uniformly labeled sample, the fibrillization of CsgA is very sensitive and highly pure monomers are required for homogeneous folding. The purification strategy was improved so that highly pure R1-labeled CsgA was produced and fibrillized homogeneously.

The peak positions in 2D ^{13}C-^{13}C solid state NMR spectra fit perfectly to the uniformly labeled sample. The signal to noise ratio and homogeneity were noticeably improved in comparison to the first samples. The linewidth was less than 1 ppm which can be consider as high resolution and is comparable to the full length CsgA spectrum. It can be concluded that the contaminations and uncontrolled fibrillization environment led to heterogeneity in the sample which could be resolved by further purification of the splicing product and fibrillization under defined conditions. Surprisingly, the general peak intensity of the segmentally labeled sample was even better than from the uniformly labeled one resulting in strong peaks even for residues which are supposed to be in a loop region. This can be explained by an even lower heterogeneity of the segmentally labeled sample caused by seeding with wild type curli while the uniformly labeled sample was produced without an addition of seeds. Another possible explanation is based on the layer specific labeling pattern which reduces the number of surrounding nuclei and potential magnetization acceptors during spin diffusion. In a tightly packed fibril the distance from a β-sheet residue Cα to the intraresidue Cγ is almost the same like to a Cα or Co in the subjacent layer. Neither the subjacent nor intermolecular β-sheets are labeled with isotopes in the R1 sample resulting in higher resolved spectra.

The ^{15}N signal to noise ratio appeared to be relatively low but sufficient for basic N-C correlation spectra which led to the almost complete sequential assignment of the R1 repeat. Cross peaks of residues Tyr51, Asp63 and Asn66 could not be identified either because of spectral overlap or intrinsic low signal caused by local flexibility.

The R2 sample showed a much lower signal to noise ratio than the R1 sample which did not allow ^{15}N-^{13}C correlations for sequential amino acid assignments. Difficulties during the purification of the splicing product resulted in remaining contaminations in sample. Relatively sharp lines were obtained in the spectrum indicating homogeneity at least for a part of the sample. Seeding with wild type curli probably induced a controlled aggregation of the full length CsgA molecules while random aggregation of contaminations disturbed the fibrillization to a certain extent. The protein pellet which was filled in the ssNMR rotor latterly contained, beside CsgA fibrils, also aggregates of intein fusion proteins. Since no peaks were detected that could not be assigned to CsgA, it can be concluded that the contaminating protein is the unlabeled intein fusion protein (IntC35). Even without any influence on the homogeneity of CsgA every contamination lowers the detected signal just by occupying space in the solid state NMR rotor.
Residues with a characteristic chemical shift could be assigned from the 2D ^{13}C-^{13}C spectra: the cross peaks of Ile71 and Ala80 were separated from R1 located Ile/Ala residues while Val82 is the only valine in CsgA R1 and R2. The sample quality could be improved by careful purification of the splicing product to obtain pure monomeric CsgA prior to fibrillization.
The structurally well characterized amyloid fibrils of the C-terminal domain of the HET-s protein from *Podospora anserina* have been used to demonstrate the power of segmental isotope labeling for solid state NMR. In this protein, residues 218-289 enclose 4 β-strands which stack up pair wise to form a left handed β-solenoid (Wasmer, Lange et al. 2008). Extraordinary sharp solid state NMR lines could be obtained from HET-s samples which led to a full assignment and a structure based on distance restraints. The assignment is available on the Biological Magnetic Resonance Bank (11028) and the structure at the Protein Data Base (2kj3). Two segmentally labeled samples were produced, labeled either on the first or the latter two β-strands.
The obtained peaks in 2D ssNMR fit exactly to a uniformly labeled sample and the published assignment demonstrating a structural identity with wild type fibrils. Just the rigid residues contribute to ssNMR and exactly those that have been assigned to b-sheets were detected in the segmentally labeled samples. Even though there is just minor spectral overlap in spectra of HET-s (218-289) more crowded regions like the Cα-CO cross peaks were nicely resolved by the labeling approach. Judged from solid state NMR signal to noise ratio and peak linewidth, good segmentally labeled samples were produced but the quality remained slightly under the expectations considering the exceptional sharp lines obtained from wild type samples. The expression yield of the IntC HET-s fusion was low which resulted in minor amounts of splicing product. The use of the Thioredoxin expression system as described for the R5 repeat of CsgA could improve the expression yield and fusion protein stability. Reversed affinity chromatography was used to purify the splicing product and the fibrillization was obtained by buffer exchange to TRIS pH 8 without GuHCl. To improve the purity possible side product from

intein cleavage should be removed by size exclusion chromatography. Additionally it has been shown that even more homogeneous samples are obtained by first desalting into a buffer at pH3 and then initiate the fibrillization by rapidly adjusting the pH to 8.
An additional challenge during structural studies of amyloid fibrils is the discrimination between intra- and intermolecular distance restraints. This issue is discussed in more detail in the next section.

5.7 UNAMBIGUOUS ASSIGNMENTS OF INTRAMOLECULAR DISTANCE RESTRAINTS WITHIN AMYLOID FIBRILS BY SEGMENTAL ISOTOPE LABELING

Amyloid fibrils are protein filaments in which b-strands from adjacent molecules form a continuous b-sheet along the fibril axis. In amyloid fibrils, where backbone and side chains of neighboring molecules arrange in close proximity, intermolecular distance restraints can appear as intense as the intramolecular ones, since the distances are similar. Structure calculations from solid state NMR data rely on distance restraints. The unambiguous assignment of long range distance restraints which define the structure is challenging in usually crowded 2D ^{13}C-^{13}C spectra. In the case of amyloid fibrils or other continuous protein arrays those restraints are either intra- or intermolecular, which have to be clearly distinguished in the assignment. Intermolecular distance restraints can be obtained by exploiting samples with mixed labeling schemes. The most straightforward way is to use 1:1 mixtures of ^{13}C and ^{15}N labeled samples and collect ^{15}N-^{13}C correlations but also mixtures of 1-^{13}C and 2-^{13}C glucose were used to assign unambiguous intermolecular distances (De Paepe, Lewandowski et al. 2011, Loquet, Sgourakis et al. 2012). Also for the elucidation of the HET-s structure a ^{15}N-^{13}C mixed sample was used for the determination of intermolecular restraints. The assignment of unambiguous intramolecular distance restraints was accomplished by sample dilution (Wasmer, Lange et al. 2008). Assuming a one dimensional linear array within the fibril, mixing of labeled and unlabeled HET-s monomers (ratio 1:2-4) under denaturing conditions prior to fibrillization results in a labeling pattern in which every labeled molecule is surrounded by unlabeled ones. The majority of distance restraints were identified as inter β-strand within one β-sheet, meaning from one layer to the other. Just very few inter β-sheets restraints could be assigned which are important to define the 3 dimensional arrangements of the β-sheets.
By layer specific segmental isotope labeling the assignment of unambiguous intramolecular distance restraints is facilitated. No further sample dilution is necessary due to the alternating labeling scheme that automatically avoids intermolecular contacts. Two dimensional ^{13}C-^{13}C spectra with long mixing times, commonly used to obtain ssNMR distance restraints, appear to be crowded since they contain sequential, medium and long range restraints. By labeling just a single transient plane with stable isotopes important information about the β-sheet arrangement can be extracted. While neither intermolecular nor inter-plane distance restraints are detectable in the CsgA R1 sample and the HET-s sample, inter-plane distances could be obtained from a CsgA R1+2 sample. Using segmental labeling for HET-s fibrils, numerous sequential, 10 medium as well as 8 long range distance restrains were unambiguously identified with a simple

2D DARR spectrum of HET-s (^{13}C 218-257). By comparing 2D DARR spectra it has been proven that the assignment of longs range restraints is highly facilitated. The peak intensity is even increased due to the reduced number of potential acceptors of magnetization during spin diffusion.

Sample dilution might be problematic on system with an intrinsic low signal to noise ratio like CsgA. Segmental isotope labeling is a promising alternative even though it is technically more challenging. Considering the R1 labeled sample, just every fifth layer in the fibrils is labeled with isotopes. Intermolecular distance restraints could just be detected in a head to head arrangement. The obtained spectra did not shown any indications for that, such as peak broadening due to equal residue contacts or interstrand distance restraints. Therefore, all distance restraints that were obtained from long mixing time ^{13}C-^{13}C spectra of R1 CsgA can be considered as intramolecular and used to define the structure of a CsgA monomer within the fibril.

5.8 Solid state NMR data gives rise to a domain swap in CsgA amyloid fibrils

Amyloid fibrils consist of b-sheets along the fibril axis. Structural differences are found in the number of β-sheets per monomer, the arrangement of β-strands in the monomer and in the intermolecular arrangement (Toyama and Weissman 2010). Highly homogeneous samples and a rigorous solid state NMR analysis are required to obtain detailed information. During this work, the protein backbone and side chains of the R1 repeat and a few residues from R2 could be assigned from segmentally labeled samples. Additionally some distance restraints were obtained for the R1 sample.

Secondary structure elements can be located based on Cα and Cβ chemical shift deviations from random coil values. The N-terminal Domain ranging from residue 21-42 did not contribute to the ssNMR spectra and is therefore flexible and not involved in secondary structure. Two β-sheets have been identified in the R1 repeat spanning from Ser43 to Tyr50 and Ser55 to Thr61 which is in agreement with the H/D exchange data (Zimmer 2011). Asn54 which was previously not clearly allocated is not involved in a β-sheet. The labeled segment of the R1 sample ranged until Leu68, which was clearly assigned and marks the first amino acid of the next β-sheet. Due to a biphasic H/D exchange rate of Asp67, the involvement into secondary structure was not entirely certain but the chemical shift assignment of Asp67 gives no indication of participation into secondary structure. Therefore Asp67 forms probably the last residue of a loop between two layers of the β-strand in the fibril.

Every residue that could be assigned from the R2 sample is involved into β-sheet. Based on the H/D exchange data they range from Asp67 to His73 and Asn77 to Gly84. Assuming a conserved fold between R1 and R2, Asn77 might not be part of the β-sheet. Ala79 has not been assigned in the solution NMR data but the solid state NMR confirms the hypothesis of β-sheet assignment.

A b-turn between Gln49 and Ser55 has been identified based on ssNMR distance restraints. This finding is in accordance with the proposed strand-turn-strand motive of CsgA in which hydrophobic residues point inwards, forming a stable amyloid core. Within the R1 repeat the hydrophobic residues Leu45, Ile47, Ala56 and Ala58 are believed to face inside the fibril. The lack of intersheet distance restraints gives rise to several structural models which are illustrated in figure 48.

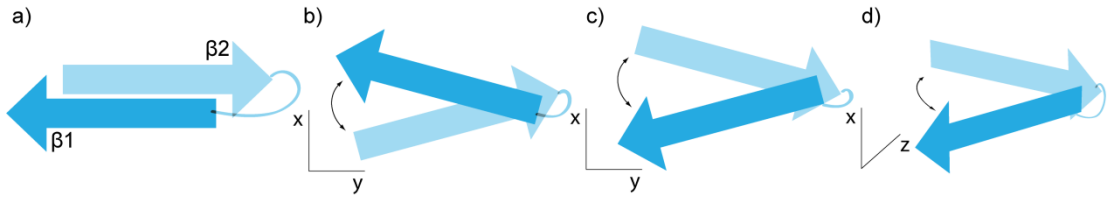


Figure 48: Models of potential β-strand orientations in CsgA fibrils.

No intersheet distance restraints, like Ile (β1) to-Ala (β2), have been detected in 2D DARR spectra but a parallel arrangement of β-sheets like shown in figure 48 a cannot be excluded so far. NMR Pulse sequences better suitable for the detection of long range restraints like PAR or CHHC have to be applied on a R1 labeled sample (Lange, Luca et al. 2002, De Paepe, Lewandowski et al. 2008). In the case of not detecting interstrand distance restraints other folds like depicted in figure 48 b,c, and e become more likely. A reasonable structural model is based on tilted b-strands giving rise to a intermolecular domain swap either by β1, which in this case forms hydrogen bonds with a β-strand of the adjacent molecule while β2 binds to strands of R2 (fig. 48b), or by β2 which in this case binds to the adjacent molecule (fig. 48c). Within these models the first one is much more realistic because the terminal β-strand residue of R1 and the initial b-strand residue of R2 would come in closer distance to each other while in the latter model those residues are further separated.

Since less is known about diversity within amyloid structures completely unexpected folds can be discussed. A theoretical alternative explanation for the absence of long range restraints is a parallel fold with a broader gap in between the β-strands (Figure 48 d). The distance would excess the possible radius for the formation of side chain hydrogen bonds and open the hydrophobic core. This fold can just be stable if either several fibrils wind around each other or other molecules are bound within the cleft. Furthermore, considering just the first two repeats, β-sheets could be composed instead of β1+β3 and β2+β4 in an arrangement as β1+β4 and β2+β3.

More investigations have to be undertaken towards a high resolution structure of CsgA. Information towards the correct model can for instance be assessed by measuring distance restraints on an improved R2 sample. According to the models b) distances from β2 to β3 should be detectable; c) distances from β1 to β4 or d) no distances from adverse β-strands would be obtained.

5.9 CONCLUSION

Amyloid fibrils of the major curli subunit CsgA were investigated by solid state NMR. Since spectra of uniformly labeled samples showed a severe signal overlap, segmental isotope labeling was applied to reduce the spectral complexity. A protein *trans* splicing based strategy for the production of R1, R1+2 and R5 labeled CsgA was established. R1 labeled samples resulted in homogeneous samples which allowed sequence specific assignments. The supposed strand-loop-strand motive could be confirmed by the secondary structure specific backbone chemical shift deviations. The lack of distance restraints gave rise to a domain swap in the CsgA fibril. Further samples need to be produced to identify unambiguous intramolecular inter β-sheet and intermolecular distance restraints.

6 OUTLOOK

In the recent study a bacterial functional amyloid system, CsgA, was investigated by solid state NMR. The requirements for calculating a protein structure based on solid state NMR data are: homogeneous samples and a decent set of NMR experiments to obtain a sequence specific resonance assignment and numerous inter- and intramolecular distance restraints. Since spectra of uniformly labeled samples did not deliver the sufficient resolution, segmental labeling was used to obtain unambiguous assignments and intramolecular distance restraints for the R1 repeat of CsgA. The missing information for a high resolution structure were intra β-sheet restraints, intramolecular inter β-sheet restraints as well as intermolecular restraints. Since the R1 turned out to be the most stable system for segmental labeling several samples were produced that combine ^{13}C labeling in the R1 repeat and a amino acid specific labeling pattern in the repeats R2-5. The labeling patterns are illustrated in figure 49.

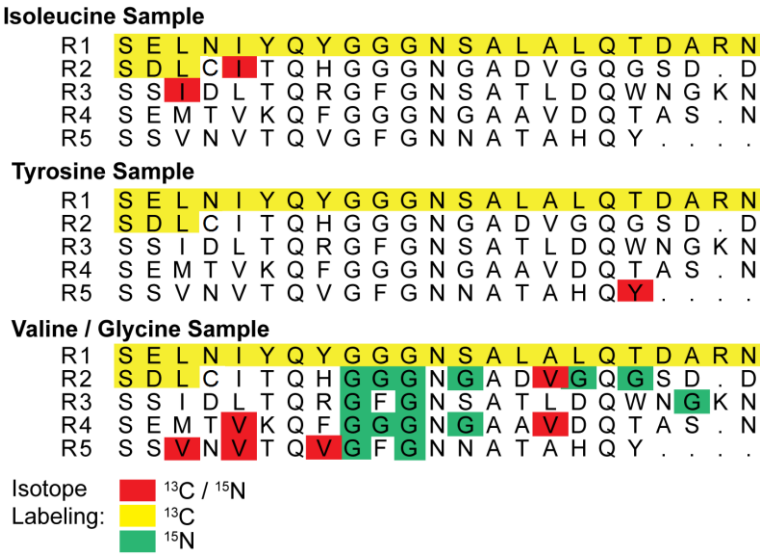


Figure 49: Segmental and amino acid specific labeled CsgA samples for the detection on unambiguous distance restraints

The CsgA repeats R2-5 contain two isoleucines, one located in R2 and the other one in R3. The isoleucine labeled sample can be exploited to obtain intra β-sheet information from R2 to R1 and maybe even from R3 to R2 either very specific by NCACX or from regular ^{13}C-^{13}C correlations. Just a single tyrosine is present in the R2-5 which is located in the very end of the R5 repeat and the tyrosine sample can therefore be used for specific intermolecular structural information. Several valines are located in the R2-5 repeats. Within those Val81 which already assigned in the R2 repeat and might provide intra β-sheet information. Furthermore three valines are located in the R5 repeat which can be used to obtain intermolecular information even though with a certain degree of ambiguity. Additionally the glycine residues are labeled with ^{15}N to provide a platform to measure NCOCX spectra and obtain unambiguous information from the two val - gly pairs in the sample.

All requirements for the production of a R 1+2 and a R5 labeled sample were set during this work. Assuming that a sequence specific assignment of the R2 and R5 repeat is feasible both sample can provide very useful information. The R 1+2 sample may provide intra β-sheet information as well as insights how the β-sheets fold on each other. The intermolecular arrangement can be scanned with a R1 / R5 mixed sample.

Taking the data of the solid state NMR and H/D-exchange experiments together it should be sufficient information to calculate a precise structure.

The structure of CsgA will provide the molecular basis for a better understanding of amyloid formation and regulation. The mechanism of curli nucleation by CsgB should be investigated as well as the influence of nucleation on the CsgA structure. Furthermore curli bind to several host proteins like fibronectin or toll like receptor 2. The structure of CsgA would be a starting point for the investigation of the molecular mechanism of the protein interaction.

7 ACKNOWLEDGEMENTS

I thank my supervisor Prof. Christiane Ritter for the opportunity to learn about solid state NMR and amyloid research in the pleasant environment of her research group.

Many thanks, to Prof. Michael Steinert for being available as second examiner and to Dr. Michael Hust for being available as chair of the examination committee.

I would like to thank Johannes Spehr for lots of productive discussions and fun. Furthermore I thank all the members of our research group MMIA namely Lichun He, Carolin Schaper, Madhu Nagaraj, Mumdooh Ahmed, Puwei Yuan, Agnes Zimmer as well as the groups SBIB, MOSB and SBAU of the Helmholtz centre for infection research. All the effort of Manfred Rohde (EM) and Anja Meier (MS) is highly appreciated.

Thanks to our collaboration partners in Berlin, especially Barth Jan van Rossum.

Thanks to the HZI Graduate School for providing a scholarship and giving me the chance to work and live in a highly international environment.

Many thanks, to all the friends from all over the world that I made. You made the time special for me and you are the main reason why I will keep Braunschweig in good memories. I hope we stay in touch. Special thanks to Sonia for the support and proofreading in the last month of my thesis.

Last but not least, I would like to thank my family. Even if I never call or visit, I always know that you are there for me when I need it, and this makes me feel calm and serene. I appreciate that a lot.

8 REFERENCES

Aguzzi, A., C. Sigurdson and M. Heikenwaelder (2008). "Molecular mechanisms of prion pathogenesis." Annual review of pathology **3**: 11-40.

Andrew, E. (1981). "Magic angle spinning." International Reviews in Physical Chemistry **1**(2): 195-224.

Anfinsen, C. (1973). "Principles that govern the folding of protein chains." Science **181**: 223-230.

Antzutkin, O. N., J. J. Balbach, R. D. Leapman, N. W. Rizzo, J. Reed and R. Tycko (2000). "Multiple quantum solid-state NMR indicates a parallel, not antiparallel, organization of beta-sheets in Alzheimer's beta-amyloid fibrils." Proc Natl Acad Sci U S A **97**(24): 13045-13050.

Austin, J. W., G. Sanders, W. W. Kay and S. K. Collinson (1998). "Thin aggregative fimbriae enhance Salmonella enteritidis biofilm formation." FEMS Microbiol Lett **162**(2): 295-301.

Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moor, J. G. Seidman, J. A. Smith and K. Struhl (2007). "Current protocols in molecular biology." John Wiley & Sons Inc.

Baldus, M., A. T. Petkova, J. Herzfeld and R. G. Griffin (1998). "Cross polarization in the tilted frame: assignment and spectral simplification in heteronuclear spin systems." Molecular Physics **95**(6): 1197-1207.

Balguerie, A., S. Dos Reis, C. Ritter, S. Chaignepain, B. Coulary-Salin, V. Forge, K. Bathany, I. Lascu, J. M. Schmitter, R. Riek and S. J. Saupe (2003). "Domain organization and structure-function relationship of the HET-s prion protein of Podospora anserina." EMBO J **22**(9): 2071-2081.

Barnhart, M. M. and M. R. Chapman (2006). "Curli biogenesis and function." Annual review of microbiology **60**: 131-147.

Bennett, A. E., C. M. Rienstra, M. Auger, K. V. Lakshmi and R. G. Griffin (1995). "Heteronuclear decoupling in rotating solids." The Journal of Chemical Physics **103**(16): 6951-6958.

Bensadoun, A. and D. Weinstein (1976). "Assay of proteins in the presence of interfering materials." Anal Biochem **70**(1): 241-250.

Berg JM, T. J., Stryer L (2002). "Biochemistry."

Bian, Z. and S. Normark (1997). "Nucleator function of CsgB for the assembly of adhesive surface organelles in Escherichia coli." EMBO J **16**(19): 5827-5836.

Bloch, F. (1946). "Nuclear Induction." Physical Review **70**(7-8): 460-474.

Bloembergen, N. (1949). "On the interaction of nuclear spins in a crystalline lattice." Physica **15**(3–4): 386-426.

Bloembergen, N., E. M. Purcell and R. V. Pound (1947). "Nuclear magnetic relaxation." Nature **160**(4066): 475.

Bockmann, A. (2008). "3D protein structures by solid-state NMR spectroscopy: ready for high resolution." Angew Chem Int Ed Engl **47**(33): 6110-6113.

Bockmann, A., C. Gardiennet, R. Verel, A. Hunkeler, A. Loquet, G. Pintacuda, L. Emsley, B. H. Meier and A. Lesage (2009). "Characterization of different water pools in solid-state NMR protein samples." J Biomol NMR **45**(3): 319-327.

Bockmann, A. and B. H. Meier (2010). "Prions: En route from structural models to structures." Prion **4**(2): 72-79.

Bonar, L., A. S. Cohen and M. M. Skinner (1969). "Characterization of the amyloid fibril as a cross-beta protein." Proc Soc Exp Biol Med **131**(4): 1373-1375.

Brunger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson and G. L. Warren (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." Acta Crystallogr D Biol Crystallogr **54**(Pt 5): 905-921.

Buchinger, E., F. L. Aachmann, A. S. Aranko, S. Valla, G. Skjak-Braek, H. Iwai and R. Wimmer (2010). "Use of protein trans-splicing to produce active and segmentally (2)H, (15)N labeled mannuronan C5-epimerase AlgE4." Protein Sci **19**(8): 1534-1543.

Busche, A. E., A. S. Aranko, M. Talebzadeh-Farooji, F. Bernhard, V. Dotsch and H. Iwai (2009). "Segmental isotopic labeling of a central domain in a multidomain protein by protein trans-splicing using only one robust DnaE intein." Angew Chem Int Ed Engl **48**(33): 6128-6131.

Castellani, F., B. van Rossum, A. Diehl, M. Schubert, K. Rehbein and H. Oschkinat (2002). "Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy." Nature **420**(6911): 98-102.

Castellani, F., B. J. van Rossum, A. Diehl, K. Rehbein and H. Oschkinat (2003). "Determination of solid-state NMR structures of proteins by means of three-dimensional 15N-13C-13C dipolar correlation spectroscopy and chemical shift analysis." Biochemistry **42**(39): 11476-11483.

Caughey, B. W., A. Dong, K. S. Bhat, D. Ernst, S. F. Hayes and W. S. Caughey (1991). "Secondary structure analysis of the scrapie-associated protein PrP 27-30 in water by infrared spectroscopy." Biochemistry **30**(31): 7672-7680.

Cavanagh, J. (2007). Protein NMR spectroscopy : principles and practice. Amsterdam ; Boston, Academic Press.

Cegelski, L., J. S. Pinkner, N. D. Hammer, C. K. Cusumano, C. S. Hung, E. Chorell, V. Aberg, J. N. Walker, P. C. Seed, F. Almqvist, M. R. Chapman and S. J. Hultgren (2009). "Small-molecule inhibitors target Escherichia coli amyloid biogenesis and biofilm formation." Nat Chem Biol **5**(12): 913-919.

Chapman, M. R., L. S. Robinson, J. S. Pinkner, R. Roth, J. Heuser, M. Hammar, S. Normark and S. J. Hultgren (2002). "Role of Escherichia coli curli operons in directing amyloid fiber formation." Science **295**(5556): 851-855.

Chen, B., K. R. Thurber, F. Shewmaker, R. B. Wickner and R. Tycko (2009). "Measurement of amyloid fibril mass-per-length by tilted-beam transmission electron microscopy." Proc Natl Acad Sci U S A **106**(34): 14339-14344.

Chen, M., M. Margittai, J. Chen and R. Langen (2007). "Investigation of alpha-synuclein fibril structure by site-directed spin labeling." The Journal of biological chemistry **282**(34): 24970-24979.

Chiba, T., Y. Hagihara, T. Higurashi, K. Hasegawa, H. Naiki and Y. Goto (2003). "Amyloid fibril formation in the context of full-length protein: effects of proline mutations on the amyloid fibril formation of beta2-microglobulin." The Journal of biological chemistry **278**(47): 47016-47024.

Chiti, F. and C. M. Dobson (2006). "Protein misfolding, functional amyloid, and human disease." Annual review of biochemistry **75**: 333-366.

Coligan, J. E. (2003). "Short Protocols in protein science. A compendium of methods from Current protocols in protein science." John Wiley & Sons Inc.

Collinson, S. K., L. Emody, K. H. Muller, T. J. Trust and W. W. Kay (1991). "Purification and characterization of thin, aggregative fimbriae from Salmonella enteritidis." J Bacteriol **173**(15): 4773-4781.

Collinson, S. K., J. M. Parker, R. S. Hodges and W. W. Kay (1999). "Structural predictions of AgfA, the insoluble fimbrial subunit of Salmonella thin aggregative fimbriae." J Mol Biol **290**(3): 741-756.

Comellas, G. and C. M. Rienstra (2013). "Protein Structure Determination by Magic-Angle Spinning Solid-State NMR, and Insights into the Formation, Structure, and Stability of Amyloid Fibrils." Annu Rev Biophys.

Cornilescu, G., F. Delaglio and A. Bax (1999). "Protein backbone angle restraints from searching a database for chemical shift and sequence homology." J Biomol NMR **13**(3): 289-302.

Coustou, V., C. Deleu, S. Saupe and J. Begueret (1997). "The protein product of the het-s heterokaryon incompatibility gene of the fungus Podospora anserina behaves as a prion analog." Proceedings of the National Academy of Sciences of the United States of America **94**(18): 9773.

De Paepe, G., J. R. Lewandowski, A. Loquet, A. Bockmann and R. G. Griffin (2008). "Proton assisted recoupling and protein structure determination." J Chem Phys **129**(24): 245101.

De Paepe, G., J. R. Lewandowski, A. Loquet, M. Eddy, S. Megy, A. Bockmann and R. G. Griffin (2011). "Heteronuclear proton assisted recoupling." The Journal of Chemical Physics **134**(9): 095101-095118.

Debelouchina, G. T., G. W. Platt, M. J. Bayro, S. E. Radford and R. G. Griffin (2010). "Magic Angle Spinning NMR Analysis of β2-Microglobulin Amyloid Fibrils in Two Distinct Morphologies." Journal of the American Chemical Society **132**(30): 10414-10423.

Dill, K. A. and H. S. Chan (1997). "From Levinthal to pathways to funnels." Nat Struct Biol **4**(1): 10-19.

Dos Reis, S., B. Coulary-Salin, V. Forge, I. Lascu, J. Begueret and S. J. Saupe (2002). "The HET-s prion protein of the filamentous fungus Podospora anserina aggregates in vitro into amyloid-like fibrils." The Journal of biological chemistry **277**(8): 5703-5706.

Dueholm, M. S., S. B. Nielsen, K. L. Hein, P. Nissen, M. Chapman, G. Christiansen, P. H. Nielsen and D. E. Otzen (2011). "Fibrillation of the major curli subunit CsgA under a wide range of conditions implies a robust design of aggregation." Biochemistry **50**(39): 8281-8290.

Eanes, E. D. and G. G. Glenner (1968). "X-ray diffraction studies on amyloid filaments." J Histochem Cytochem **16**(11): 673-677.

Edman, P. and G. Begg (1967). "A protein sequenator." Eur J Biochem **1**(1): 80-91.

Eisenberg, D. and M. Jucker (2012). "The amyloid state of proteins in human diseases." Cell **148**(6): 1188-1203.

Englander, S. W. and N. R. Kallenbach (1983). "Hydrogen exchange and structural dynamics of proteins and nucleic acids." Quarterly reviews of biophysics **16**(04): 521-655.

Ernst, R., G. Bodenhausen and A. Wokaun (1987). "Principles of nuclear magnetic resonance in one and two dimensions." Oxford: Clarendon Press.

Fandrich, M., J. Meinhardt and N. Grigorieff (2009). "Structural polymorphism of Alzheimer Abeta and other amyloid fibrils." Prion **3**(2): 89-93.

Fitzpatrick, A. W., G. T. Debelouchina, M. J. Bayro, D. K. Clare, M. A. Caporini, V. S. Bajaj, C. P. Jaroniec, L. Wang, V. Ladizhansky, S. A. Muller, C. E. Macphee, C. A. Waudby, H. R. Mott, A. De Simone, T. P. Knowles, H. R. Saibil, M. Vendruscolo, E. V. Orlova, R. G. Griffin and C. M. Dobson (2013). "Atomic structure and hierarchical assembly of a cross-beta amyloid fibril." Proc Natl Acad Sci U S A **110**(14): 5468-5473.

Fowler, D. M., A. V. Koulov, C. Alory-Jost, M. S. Marks, W. E. Balch and J. W. Kelly (2006). "Functional amyloid formation within mammalian tissue." PLoS Biol **4**(1): e6.

Fowler, D. M., A. V. Koulov, W. E. Balch and J. W. Kelly (2007). "Functional amyloid - from bacteria to humans." Trends in biochemical sciences **32**(5): 217-224.

Fung, B. M., A. K. Khitrin and K. Ermolaev (2000). "An improved broadband decoupling sequence for liquid crystals and solids." J Magn Reson **142**(1): 97-101.

Gerstel, U., C. Park and U. Romling (2003). "Complex regulation of csgD promoter activity by global regulatory proteins." Molecular microbiology **49**(3): 639-654.

Glenner, G. G., P. Cuatrecasas, C. Isersky, H. A. Bladen and E. D. Eanes (1969). "Physical and chemical properties of amyloid fibers. II. Isolation of a unique protein constituting the major component from human splenic amyloid fibril concentrates." J Histochem Cytochem **17**(12): 769-780.

Goldberg, A. L. (2003). "Protein degradation and protection against misfolded or damaged proteins." Nature **426**(6968): 895-899.

Gophna, U., M. Barlev, R. Seijffers, T. A. Oelschlager, J. Hacker and E. Z. Ron (2001). "Curli fibers mediate internalization of Escherichia coli by eukaryotic cells." Infect Immun **69**(4): 2659-2665.

Greenwald, J., C. Buhtz, C. Ritter, W. Kwiatkowski, S. Choe, M.-L. Maddelein, F. Ness, S. Cescau, A. Soragni, D. Leitz, S. J. Saupe and R. Riek (2010). "The Mechanism of Prion Inhibition by HET-S." Molecular Cell **38**(6): 889-899.

Greenwald, J. and R. Riek (2010). "Biology of amyloid: structure, function, and regulation." Structure **18**(10): 1244-1260.

Griffin, R. G. (1998). "Dipolar recoupling in MAS spectra of biological solids." Nature structural biology **5 Suppl**: 508-512.

Guntert, P., C. Mumenthaler and K. Wuthrich (1997). "Torsion angle dynamics for NMR structure calculation with the new program DYANA." J Mol Biol **273**(1): 283-298.

Habenstein, B., C. Wasmer, L. Bousset, Y. Sourigues, A. Schutz, A. Loquet, B. H. Meier, R. Melki and A. Bockmann (2011). "Extensive de novo solid-state NMR assignments of the 33 kDa C-terminal domain of the Ure2 prion." J Biomol NMR **51**(3): 235-243.

Hall, T. M. T., J. A. Porter, K. E. Young, E. V. Koonin, P. A. Beachy and D. J. Leahy (1997). "Crystal Structure of a Hedgehog Autoprocessing Domain: Homology between Hedgehog and Self-Splicing Proteins." Cell **91**(1): 85-97.

Hammar, M., A. Arnqvist, Z. Bian, A. Olsen and S. Normark (1995). "Expression of two csg operons is required for production of fibronectin- and congo red-binding curli polymers in Escherichia coli K-12." Molecular microbiology **18**(4): 661-670.

Hammer, N. D., J. C. Schmidt and M. R. Chapman (2007). "The curli nucleator protein, CsgB, contains an amyloidogenic domain that directs CsgA polymerization." Proc Natl Acad Sci U S A **104**(30): 12494-12499.

Hammer, N. D., X. Wang, B. A. McGuffie and M. R. Chapman (2008). "Amyloids: friend or foe?" J Alzheimers Dis **13**(4): 407-419.

Harper, J. D. and P. T. Lansbury, Jr. (1997). "Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins." Annu Rev Biochem **66**: 385-407.

Hartl, F. U., A. Bracher and M. Hayer-Hartl (2011). "Molecular chaperones in protein folding and proteostasis." Nature **475**(7356): 324-332.

Hartl, F. U. and M. Hayer-Hartl (2009). "Converging concepts of protein folding in vitro and in vivo." Nature Structural & Molecular Biology **16**(6): 574-581.

Hartmann, S. R. and E. L. Hahn (1962). "Nuclear Double Resonance in the Rotating Frame." Physical Review **128**(5): 2042-2053.

Heise, H., W. Hoyer, S. Becker, O. C. Andronesi, D. Riedel and M. Baldus (2005). "Molecular-level secondary structure, polymorphism, and dynamics of full-length alpha-synuclein fibrils studied by solid-state NMR." Proceedings of the National Academy of Sciences of the United States of America **102**(44): 15871-15876.

Herczenik, E. and M. F. B. G. Gebbink (2008). "Molecular and cellular aspects of protein misfolding and disease." The FASEB Journal **22**(7): 2115.

Higman, V., J. Flinders, M. Hiller, S. Jehle, S. Markovic, S. Fiedler, B.-J. Rossum and H. Oschkinat (2009). "Assigning large proteins in the solid state: a MAS NMR resonance assignment strategy using selectively and extensively 13C-labelled proteins." Journal of Biomolecular NMR **44**(4): 245-260.

Hong, M., Y. Zhang and F. Hu (2012). "Membrane protein structure and dynamics from NMR spectroscopy." Annu Rev Phys Chem **63**: 1-24.

Hoshino, M., H. Katou, Y. Hagihara, K. Hasegawa, H. Naiki and Y. Goto (2002). "Mapping the core of the β2-microglobulin amyloid fibril by H/D exchange." Nature structural biology **9**(5): 332-336.

Howie, A. J. and D. B. Brewer (2009). "Optical properties of amyloid stained by Congo red: history and mechanisms." Micron **40**(3): 285-301.

Hvidt, A. and K. Linderstrøm-Lang (1954). "Exchange of hydrogen atoms in insulin with deuterium atoms in aqueous solutions." Biochimica et biophysica acta **14**(4): 574.

Ivankov, D. N. and A. V. Finkelstein (2004). "Prediction of protein folding rates from the amino acid sequence-predicted secondary structure." Proc Natl Acad Sci U S A **101**(24): 8942-8944.

Iwai, H., S. Zuger, J. Jin and P. H. Tam (2006). "Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostoc punctiforme." FEBS Lett **580**(7): 1853-1858.

Jarrett, J. T. and P. T. Lansbury, Jr. (1993). "Seeding "one-dimensional crystallization" of amyloid: a pathogenic mechanism in Alzheimer's disease and scrapie?" Cell **73**(6): 1055-1058.

Jimenez, J. L., J. I. Guijarro, E. Orlova, J. Zurdo, C. M. Dobson, M. Sunde and H. R. Saibil (1999). "Cryo-electron microscopy structure of an SH3 amyloid fibril and model of the molecular packing." The EMBO journal **18**(4): 815-821.

Kang, J., H. G. Lemaire, A. Unterbeck, J. M. Salbaum, C. L. Masters, K. H. Grzeschik, G. Multhaup, K. Beyreuther and B. Muller-Hill (1987). "The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor." Nature **325**(6106): 733-736.

Kent, S. B. (2009). "Total chemical synthesis of proteins." Chem Soc Rev **38**(2): 338-351.

Kheterpal, I., K. D. Cook and R. Wetzel (2006). "Hydrogen/deuterium exchange mass spectrometry analysis of protein aggregates." Methods in enzymology **413**: 140-166.

Kikuchi, T., Y. Mizunoe, A. Takade, S. Naito and S. Yoshida (2005). "Curli fibers are required for development of biofilm architecture in Escherichia coli K-12 and enhance bacterial adherence to human uroepithelial cells." Microbiol Immunol **49**(9): 875-884.

Knowles, T. P., J. F. Smith, A. Craig, C. M. Dobson and M. E. Welland (2006). "Spatial persistence of angular correlations in amyloid fibrils." Phys Rev Lett **96**(23): 238301.

Krishnan, R. and S. L. Lindquist (2005). "Structural insights into a yeast prion illuminate nucleation and strain diversity." Nature **435**(7043): 765-772.

Lange, A., S. Luca and M. Baldus (2002). "Structural Constraints from Proton-Mediated Rare-Spin Correlation Spectroscopy in Rotating Solids†." Journal of the American Chemical Society **124**(33): 9704-9705.

Lansbury, P. T., Jr., P. R. Costa, J. M. Griffiths, E. J. Simon, M. Auger, K. J. Halverson, D. A. Kocisko, Z. S. Hendsch, T. T. Ashburn, R. G. Spencer and et al. (1995). "Structural model for the beta-amyloid fibril based on interstrand alignment of an antiparallel-sheet comprising a C-terminal peptide." Nat Struct Biol **2**(11): 990-998.

LeMaster, D. M. and D. M. Kushlan (1996). "Dynamical Mapping of E. coli Thioredoxin via 13C NMR Relaxation Analysis." Journal of the American Chemical Society **118**(39): 9255-9264.

LeVine, H., 3rd (1999). "Quantification of beta-sheet amyloid fibril structures with thioflavin T." Methods Enzymol **309**: 274-284.

Levitt, M. H. (2008). Spin dynamics : basics of nuclear magnetic resonance. Chichester, England ; Hoboken, NJ, John Wiley & Sons.

Li, Y., D. A. Berthold, H. L. Frericks, R. B. Gennis and C. M. Rienstra (2007). "Partial 13C and 15N Chemical-Shift Assignments of the Disulfide-Bond-Forming Enzyme DsbB by 3D Magic-Angle Spinning NMR Spectroscopy." ChemBioChem **8**(4): 434-442.

Lockless, S. W. and T. W. Muir (2009). "Traceless protein splicing utilizing evolved split inteins." Proc Natl Acad Sci U S A **106**(27): 10999-11004.

Lopez del Amo, J. M., M. Schmidt, U. Fink, M. Dasari, M. Fandrich and B. Reif (2012). "An asymmetric dimer as the basic subunit in Alzheimer's disease amyloid beta fibrils." Angew Chem Int Ed Engl **51**(25): 6136-6139.

Loquet, A., K. Giller, S. Becker and A. Lange (2010). "Supramolecular Interactions Probed by 13C−13C Solid-State NMR Spectroscopy." Journal of the American Chemical Society **132**(43): 15164-15166.

Loquet, A., N. G. Sgourakis, R. Gupta, K. Giller, D. Riedel, C. Goosmann, C. Griesinger, M. Kolbe, D. Baker, S. Becker and A. Lange (2012). "Atomic model of the type III secretion system needle." Nature **486**(7402): 276-279.

Luhrs, T., C. Ritter, M. Adrian, D. Riek-Loher, B. Bohrmann, H. Dobeli, D. Schubert and R. Riek (2005). "3D structure of Alzheimer's amyloid-beta(1-42) fibrils." Proc Natl Acad Sci U S A **102**(48): 17342-17347.

Maddelein, M. L., S. Dos Reis, S. Duvezin-Caubet, B. Coulary-Salin and S. J. Saupe (2002). "Amyloid aggregates of the HET-s prion protein are infectious." Proc Natl Acad Sci U S A **99**(11): 7402-7407.

Manolikas, T., T. Herrmann and B. H. Meier (2008). "Protein Structure Determination from 13C Spin-Diffusion Solid-State NMR Spectroscopy." Journal of the American Chemical Society **130**(12): 3959-3966.

Markley, J. L., A. Bax, Y. Arata, C. W. Hilbers, R. Kaptein, B. D. Sykes, P. E. Wright and K. Wuthrich (1998). "Recommendations for the presentation of NMR structures of proteins and nucleic acids--IUPAC-IUBMB-IUPAB Inter-Union Task Group on the standardization of data bases of protein and nucleic acid structures determined by NMR spectroscopy." Eur J Biochem **256**(1): 1-15.

McIntosh, L. P. and F. W. Dahlquist (1990). "Biosynthetic Incorporation of 15N and 13C for Assignment and Interpretation of Nuclear Magnetic Resonance Spectra of Proteins." Quarterly Reviews of Biophysics **23**(01): 1-38.

Mehring, M. (1983). "Principles of High Resolution NMR in Solids." Springer Verlag.

Meinhardt, J., C. Sachse, P. Hortschansky, N. Grigorieff and M. Fandrich (2009). "Aβ(1-40) fibril polymorphism implies diverse interaction patterns in amyloid fibrils." Journal of molecular biology **386**(3): 869-877.

Mootz, H. D. (2009). "Split inteins as versatile tools for protein semisynthesis." Chembiochem **10**(16): 2579-2589.

Muir, T. W., D. Sondhi and P. A. Cole (1998). "Expressed protein ligation: a general method for protein engineering." Proc Natl Acad Sci U S A **95**(12): 6705-6710.

Muona, M., A. S. Aranko, V. Raulinaitis and H. Iwai (2010). "Segmental isotopic labeling of multi-domain and fusion proteins by protein trans-splicing in vivo and in vitro." Nat Protoc **5**(3): 574-587.

Muralidharan, V. and T. W. Muir (2006). "Protein ligation: an enabling technology for the biophysical analysis of proteins." Nat Meth **3**(6): 429-438.

Näslund, J., A. Schierhorn, U. Hellman, L. Lannfelt, A. D. Roses, L. O. Tjernberg, J. Silberring, S. E. Gandy, B. Winblad and P. Greengard (1994). "Relative abundance of Alzheimer A beta amyloid peptide variants in Alzheimer disease and normal aging." Proceedings of the National Academy of Sciences **91**(18): 8378-8382.

Nelson, R., M. R. Sawaya, M. Balbirnie, A. O. Madsen, C. Riekel, R. Grothe and D. Eisenberg (2005). "Structure of the cross-beta spine of amyloid-like fibrils." Nature **435**(7043): 773-778.

Nenninger, A. A., L. S. Robinson, N. D. Hammer, E. A. Epstein, M. P. Badtke, S. J. Hultgren and M. R. Chapman (2011). "CsgE is a curli secretion specificity factor that prevents amyloid fibre aggregation." Molecular microbiology.

Nenninger, A. A., L. S. Robinson and S. J. Hultgren (2009). "Localized and efficient curli nucleation requires the chaperone-like amyloid assembly protein CsgF." Proceedings of the National Academy of Sciences of the United States of America **106**(3): 900-905.

Oeemig, J. S., A. S. Aranko, J. Djupsjobacka, K. Heinamaki and H. Iwai (2009). "Solution structure of DnaE intein from Nostoc punctiforme: structural basis for the design of a new split intein suitable for site-specific chemical modification." FEBS Lett **583**(9): 1451-1456.

Olsen, A., A. Jonsson and S. Normark (1989). "Fibronectin binding mediated by a novel class of surface organelles on Escherichia coli." Nature **338**(6217): 652-655.

Orrú, C. and B. Caughey (2011). Prion Seeded Conversion and Amplification Assays. Prion Proteins. J. Tatzelt, Springer Berlin Heidelberg. **305:** 121-133.

Paravastu, A. K., R. D. Leapman, W. M. Yau and R. Tycko (2008). "Molecular structural basis for polymorphism in Alzheimer's beta-amyloid fibrils." Proc Natl Acad Sci U S A **105**(47): 18349-18354.

Pauli, J., M. Baldus, B. van Rossum, H. de Groot and H. Oschkinat (2001). "Backbone and Side-Chain 13C and 15N Signal Assignments of the α-Spectrin SH3 Domain by Magic Angle Spinning Solid-State NMR at 17.6 Tesla." ChemBioChem **2**(4): 272-281.

Pauling, L. and R. B. Corey (1951). "Configurations of Polypeptide Chains With Favored Orientations Around Single Bonds: Two New Pleated Sheets." Proc Natl Acad Sci U S A **37**(11): 729-740.

Pauling, L., R. B. Corey and H. R. Branson (1951). "The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain." Proc Natl Acad Sci U S A **37**(4): 205-211.

Petkova, A. T., Y. Ishii, J. J. Balbach, O. N. Antzutkin, R. D. Leapman, F. Delaglio and R. Tycko (2002). "A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR." Proceedings of the National Academy of Sciences of the United States of America **99**(26): 16742-16747.

Petkova, A. T., R. D. Leapman, Z. Guo, W. M. Yau, M. P. Mattson and R. Tycko (2005). "Self-propagating, molecular-level polymorphism in Alzheimer's beta-amyloid fibrils." Science **307**(5707): 262-265.

Petkova, A. T., W. M. Yau and R. Tycko (2006). "Experimental constraints on quaternary structure in Alzheimer's beta-amyloid fibrils." Biochemistry **45**(2): 498-512.

Pines, A., M. G. Gibby and J. S. Waugh (1973). "Proton-enhanced NMR of dilute spins in solids." The Journal of Chemical Physics **59**(2): 569-590.

Prusiner, S. B. (1982). "Novel proteinaceous infectious particles cause scrapie." Science **216**(4542): 136-144.

Renault, M., A. Cukkemane and M. Baldus (2010). "Solid-state NMR spectroscopy on complex biomolecules." Angew Chem Int Ed Engl **49**(45): 8346-8357.

Ritter, C., M. L. Maddelein, A. B. Siemer, T. Luhrs, M. Ernst, B. H. Meier, S. J. Saupe and R. Riek (2005). "Correlation of structural elements and infectivity of the HET-s prion." Nature **435**(7043): 844-848.

Robinson, L. S., E. M. Ashman, S. J. Hultgren and M. R. Chapman (2006). "Secretion of curli fibre subunits is mediated by the outer membrane-localized CsgG protein." Mol Microbiol **59**(3): 870-881.

Romling, U., Z. Bian, M. Hammar, W. D. Sierralta and S. Normark (1998). "Curli fibers are highly conserved between Salmonella typhimurium and Escherichia coli with respect to operon structure and regulation." J Bacteriol **180**(3): 722-731.

Ross, E. D., H. K. Edskes, M. J. Terry and R. B. Wickner (2005). "Primary sequence independence for prion formation." Proceedings of the National Academy of Sciences of the United States of America **102**(36): 12825-12830.

Sachse, C., M. Fandrich and N. Grigorieff (2008). "Paired beta-sheet structure of an Aβ(1-40) amyloid fibril revealed by electron microscopy." Proceedings of the National Academy of Sciences of the United States of America **105**(21): 7462-7466.

Sambrook, J. and D. W. Russell (2000). "Molecular cloning - a laboratory manual." Cold Spring Harbor Laboratory Press.

Saupe, S. J. (2007). "A short history of small s: a prion of the fungus Podospora anserina." Prion **1**(2): 110-115.

Schmidt, M., C. Sachse, W. Richter, C. Xu, M. Fandrich and N. Grigorieff (2009). "Comparison of Alzheimer Aβ(1-40) and Aβ(1-42) amyloid fibrils reveals similar protofilament structures."

Proceedings of the National Academy of Sciences of the United States of America **106**(47): 19813-19818.

Schuetz, A., C. Wasmer, B. Habenstein, R. Verel, J. Greenwald, R. Riek, A. Bockmann and B. H. Meier (2010). "Protocols for the sequential solid-state NMR spectroscopic assignment of a uniformly labeled 25 kDa protein: HET-s(1-227)." Chembiochem : a European journal of chemical biology **11**(11): 1543-1551.

Schwieters, C. D., J. J. Kuszewski, N. Tjandra and G. M. Clore (2003). "The Xplor-NIH NMR molecular structure determination package." J Magn Reson **160**(1): 65-73.

Serdyuk, I. N., N. R. Zaccai and J. Zaccai (2007). "Methods in Molecular Biophysics." Cambridge University Press.

Shen, Y., O. Lange, F. Delaglio, P. Rossi, J. M. Aramini, G. Liu, A. Eletsky, Y. Wu, K. K. Singarapu, A. Lemak, A. Ignatchenko, C. H. Arrowsmith, T. Szyperski, G. T. Montelione, D. Baker and A. Bax (2008). "Consistent blind protein structure generation from NMR chemical shift data." Proceedings of the National Academy of Sciences **105**(12): 4685-4690.

Shewmaker, F., R. P. McGlinchey, K. R. Thurber, P. McPhie, F. Dyda, R. Tycko and R. B. Wickner (2009). "The functional curli amyloid is not based on in-register parallel beta-sheet structure." J Biol Chem **284**(37): 25065-25076.

Shirahama, T. and A. S. Cohen (1965). "Structure of amyloid fibrils after negative staining and high-resolution electron microscopy." Nature **206**(985): 737-738.

Siemer, A. B., C. Ritter, M. Ernst, R. Riek and B. H. Meier (2005). "High-resolution solid-state NMR spectroscopy of the prion protein HET-s in its amyloid conformation." Angew Chem Int Ed Engl **44**(16): 2441-2444.

Sipe, J. D. and A. S. Cohen (2000). "Review: history of the amyloid fibril." J Struct Biol **130**(2-3): 88-98.

Szeverenyi, N. M., M. J. Sullivan and G. E. Maciel (1982). "Observation of spin exchange by two-dimensional fourier transform 13C cross polarization-magic-angle spinning." Journal of Magnetic Resonance (1969) **47**(3): 462-475.

Takegoshi, K., S. Nakamura and T. Terao (2001). "13C–1H dipolar-assisted rotational resonance in magic-angle spinning NMR." Chemical Physics Letters **344**(5–6): 631-637.

Telenti, A., M. Southworth, F. Alcaide, S. Daugelat, W. R. Jacobs, Jr. and F. B. Perler (1997). "The Mycobacterium xenopi GyrA protein splicing element: characterization of a minimal intein." J Bacteriol **179**(20): 6378-6382.

Termine, J. D., E. D. Eanes, D. Ein and G. G. Glenner (1972). "Infrared spectroscopy of human amyloid fibrils and immunoglobulin proteins." Biopolymers **11**(5): 1103-1113.

Torok, M., S. Milton, R. Kayed, P. Wu, T. McIntire, C. G. Glabe and R. Langen (2002). "Structural and dynamic features of Alzheimer's Aβ peptide in amyloid fibrils studied by site-directed spin labeling." The Journal of biological chemistry **277**(43): 40810-40815.

Toyama, B. and J. Weissman (2010). "Amyloid Structure: Conformational Diversity and Consequences." Annual review of biochemistry.

Tükel, Ç., J. H. Nishimori, R. P. Wilson, M. G. Winter, A. M. Keestra, J. P. M. Van Putten and A. J. Bäumler (2010). "Toll-like receptors 1 and 2 cooperatively mediate immune responses to curli, a common amyloid from enterobacterial biofilms." Cellular microbiology **12**(10): 1495-1505.

Tukel, C., M. Raffatellu, A. D. Humphries, R. P. Wilson, H. L. Andrews-Polymenis, T. Gull, J. F. Figueiredo, M. H. Wong, K. S. Michelsen, M. Akcelik, L. G. Adams and A. J. Baumler (2005).

"CsgA is a pathogen-associated molecular pattern of Salmonella enterica serotype Typhimurium that is recognized by Toll-like receptor 2." Molecular microbiology **58**(1): 289-304.

Tycko, R. and R. B. Wickner (2013). "Molecular Structures of Amyloid and Prion Fibrils: Consensus versus Controversy." Acc Chem Res.

Uversky, V. and A. Fink (2006). "Protein Misfolding, Aggregation and Conformational Diseases Vols. 1." Kluwer Academic/Plenum.

Van Houdt, R. and C. W. Michiels (2005). "Role of bacterial cell surface structures in Escherichia coli biofilm formation." Research in microbiology **156**(5-6): 626-633.

Van Melckebeke, H., C. Wasmer, A. Lange, E. Ab, A. Loquet, A. Bockmann and B. H. Meier (2010). "Atomic-resolution three-dimensional structure of HET-s(218-289) amyloid fibrils by solid-state NMR spectroscopy." Journal of the American Chemical Society **132**(39): 13765-13775.

Vassar, P. S. and C. F. Culling (1959). "Fluorescent stains, with special reference to amyloid and connective tissues." Arch Pathol **68**: 487-498.

Vila-Perello, M. and T. W. Muir (2010). "Biological applications of protein splicing." Cell **143**(2): 191-200.

Volkmann, G. and H. Iwai (2010). "Protein trans-splicing and its use in structural biology: opportunities and limitations." Mol Biosyst **6**(11): 2110-2121.

Vranken, W. F., W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides and E. D. Laue (2005). "The CCPN data model for NMR spectroscopy: development of a software pipeline." Proteins **59**(4): 687-696.

Wang, X., D. R. Smith, J. W. Jones and M. R. Chapman (2007). "In vitro polymerization of a functional Escherichia coli amyloid protein." J Biol Chem **282**(6): 3713-3719.

Wang, X., Y. Zhou, J. J. Ren, N. D. Hammer and M. R. Chapman (2010). "Gatekeeper residues in the major curlin subunit modulate bacterial amyloid fiber biogenesis." Proceedings of the National Academy of Sciences of the United States of America **107**(1): 163-168.

Wasmer, C., A. Lange, H. Van Melckebeke, A. B. Siemer, R. Riek and B. H. Meier (2008). "Amyloid fibrils of the HET-s(218-289) prion form a β-solenoid with a triangular hydrophobic core." Science **319**(5869): 1523-1526.

White, A. P., S. K. Collinson, P. A. Banser, D. L. Gibson, M. Paetzel, N. C. Strynadka and W. W. Kay (2001). "Structure and characterization of AgfB from Salmonella enteritidis thin aggregative fimbriae." J Mol Biol **311**(4): 735-749.

Whittemore, N. A., R. Mishra, I. Kheterpal, A. D. Williams, R. Wetzel and E. H. Serpersu (2005). "Hydrogen-deuterium (H/D) exchange mapping of Abeta 1-40 amyloid fibril secondary structure using nuclear magnetic resonance spectroscopy." Biochemistry **44**(11): 4434-4441.

Wickner, R. B., H. K. Edskes, M.-L. Maddelein, K. L. Taylor and H. Moriyama (1999). "Prions of Yeast and Fungi." Journal of Biological Chemistry **274**(2): 555-558.

Williams, A. D., E. Portelius, I. Kheterpal, J. T. Guo, K. D. Cook, Y. Xu and R. Wetzel (2004). "Mapping Aβ amyloid fibril secondary structure using scanning proline mutagenesis." Journal of molecular biology **335**(3): 833-842.

Wu, H., Z. Hu and X.-Q. Liu (1998). "Protein trans-splicing by a split intein encoded in a split DnaE gene of Synechocystis sp. PCC6803." Proceedings of the National Academy of Sciences **95**(16): 9226-9231.

Wüthrich, K. (1986). NMR of proteins and nucleic acids. New York, Wiley.

Xu, M.-Q., M. W. Southworth, F. B. Mersha, L. J. Hornstra and F. B. Perler (1993). "In vitro protein splicing of purified precursor and the identification of a branched intermediate." Cell **75**(7): 1371-1377.

Yagi, H., T. Tsujimoto, T. Yamazaki, M. Yoshida and H. Akutsu (2004). "Conformational Change of H+-ATPase β Monomer Revealed on Segmental Isotope Labeling NMR Spectroscopy." Journal of the American Chemical Society **126**(50): 16632-16638.

Yamazaki, T., T. Otomo, N. Oda, Y. Kyogoku, K. Uegaki, N. Ito, Y. Ishino and H. Nakamura (1998). "Segmental Isotope Labeling for Protein NMR Using Peptide Splicing." Journal of the American Chemical Society **120**(22): 5591-5592.

Zeeman, P. (1897). "The effect of magnetisation on the nature of light emitted by a substance." Nature **55**(1424): 347-347.

Zhang, R., X. Hu, H. Khant, S. J. Ludtke, W. Chiu, M. F. Schmid, C. Frieden and J. M. Lee (2009). "Interprotofilament interactions between Alzheimer's Aβ1-42 peptides in amyloid fibrils revealed by cryoEM." Proceedings of the National Academy of Sciences of the United States of America **106**(12): 4653-4658.

Zimmer, A. (2011). "Strukturelle Charakterisierung von Amyloiden aus Pilzen und Bakterien mittels Wasserstoff/Deuterium Austausch NMR-Spektroskopie." http://www.digibib.tu-bs.de/?docid=00042003.

9 APPENDIX

a) Amino acid sequences

CsgAR1-IntN-His

MGVVPQYGGG GNHGGGGNNS GPNSELNIYQ YGGGNSALAL QTDARNSDLC LSYETEILTV EYGLLPIGKI VEKRIECTVY SVDNNGNIYT QPVAQWHDRG EQEVFEYCLE DGSLIRATKD HKFMTVDGQM LPIDEIFERE LDLMRVDNLP NLEHHHHHH

CsgAR1+2-IntN-His

MGVVPQYGGG GNHGGGGNNS GPNSELNIYQ YGGGNSALAL QTDARNSDLT ITQHGGGNGA DVGQGSDDSC LSYETEILTV EYGLLPIGKI VEKRIECTVY SVDNNGNIYT QPVAQWHDRG EQEVFEYCLE DGSLIRATKD HKFMTVDGQM LPIDEIFERE LDLMRVDNLP NLEHHHHHH

CsgAR1-4-IntN-His

MGVVPQYGGG GNHGGGGNNS GPNSELNIYQ YGGGNSALAL QTDARNSDLT ITQHGGGNGA DVGQGSDDSS IDLTQRGFGN SATLDQWNGK NSEMTVKQFG GGNGAAVDQT ACLSYETEIL TVEYGLLPIG KIVEKRIECT VYSVDNNGNI YTQPVAQWHD RGEQEVFEYC LEDGSLIRAT KDHKFMTVDG QMLPIDEIFE RELDLMRVDN LPNLEHHHHH H

His-IntC-CsgAR2-5

HHHHHHIKIA TRKYLGKQNV YDIGVERDHN FALKNGFIAS NITQHGGGNG ADVGQGSDDS SIDLTQRGFG NSATLDQWNG KNSEMTVKQF GGGNGAAVDQ TASNSSVNVT QVGFGNNATA HQY

His-IntC-CsgAR3-5

MHHHHHHHIKI ATRKYLGKQN VYDIGVERDH NFALKNGFIA SNCIDLTQRG FGNSATLDQW NGKNSEMTVK QFGGGNGAAV DQTASNSSVN VTQVGFGNNA TAHQY

His-IntC-CsgAR5

MHHHHHHHIKI ATRKYLGKQN VYDIGVERDH NFALKNGFIA SNCNSSVNVT QVGFGNNATA HQY

Trx-IntC-R5

MSDKIIHLTD DSFDTDVLKA DGAILVDFWA EWCGPCKMIA PILDEIADEY QGKLTVAKLN IDQNPGTAPK YGIRGIPTLL LFKNGEVAAT KVGALSKGQL KEFLDANLAG SGSGHMHHHH HHSSGENLYF QGAMIKIATR KYLGKQNVYD IGVERDHNFA LKNGFIASNC NSSVNVTQVG FGNNATAHQY

His-IntC-HETs257-289

MHHHHHHHIKI ATRKYLGKQN VYDIGVERDH NFALKNGFIA SNCDQTTNSV ETVVGKGESR VLIGNEYGGK GFWDN

HETs217-256-IntN-His

MKIDAIVGRN SAKDIRTEER ARVQLGNVVT AAALHGGIRI CLSYETEILT VEYGLLPIGK IVEKRIECTV YSVDNNGNIY TQPVAQWHDR GEQEVFEYCL EDGSLIRATK DHKFMTVDGQ MLPIDEIFER ELDLMRVDNL PNLEHHHHHH

b) List of figures

c) List of tables